THE DIAGNOSTIC/ PROGNOSTIC VALUE OF NEONATAL FINDINGS FOR PREDICTING CHILDHOOD AND ADULT MORBIDITY: SYSTEMATIC REVIEWS, META-ANALYSIS AND DECISION ANALYTIC MODELLING

By Dr Gemma Louise Malin

A thesis submitted to the University of Birmingham for the degree of:

DOCTOR OF MEDICINE

College of Medical and Dental Sciences

The University of Birmingham

October 2012

SYNOPSIS

THIS THESIS SYSTEMATICALLY REVIEWED THE PROGNOSTIC ASSOCIATION AND PREDICTIVE ABILITY OF NEONATAL TESTS FOR SHORT AND LONG TERM ADVERSE OUTCOMES, AND PERFORMED DECISION-ANALYTIC MODELLING TO EXPLORE THE IMPACT OF VARYING TEST THRESHOLD ON THE COST-EFFECTIVENESS OF UMBILICAL CORD PH TESTING AND NEONATAL THERAPEUTIC HYPOTHERMIA TO PREVENT CEREBRAL PALSY.

ABSTRACT

Events in utero have been linked with a variety of diseases throughout life, from the neonatal period to adulthood, however there is a lack of consensus regarding the ability of neonatal tests to predict these outcomes. Systematic reviews and meta-analyses were performed, assessing umbilical cord pH and base excess at birth, standards of low birth weight, and the Apgar score, including a total of 218 papers and 26704980 individuals. The prognostic association and predictive accuracy of these tests for adverse outcomes, including neonatal mortality and morbidity, childhood morbidity including cerebral palsy, and where possible adult outcomes, were determined. A decision-analytic model based analysis assessed the cost-effectiveness of varying the umbilical cord pH threshold, and treatment with neonatal hypothermia. This thesis determined that all of the tests examined had a strong association with neonatal mortality, and a significant but smaller association with neonatal morbidity and childhood cerebral palsy. In general, where the association was strong, tests had a high specificity and positive likelihood ratio for adverse outcome, but poor sensitivity and negative likelihood ratio, indicating that negative tests do not reduce the risk. The cost effectiveness analysis showed that the threshold of pH used in current practice to recommend neonatal hypothermia is more effective and less costly than a higher threshold.

EXECUTIVE SUMMARY

Background

Events in utero have been linked with a variety of diseases throughout life, from the neonatal period to adulthood. Tests performed immediately following delivery of a baby not only summarise the condition at that moment, but may reflect events in utero and during labour. A variety of tests may be performed, but practice varies widely and there is a lack of consensus regarding how the results may be interpreted and used to predict adverse outcome or indicate the need for intervention. There is a wealth of literature linking neonatal tests with mortality and morbidity throughout the life course, but results are inconsistent and there is a lack of clear collated information summarising these findings.

This thesis aimed to evaluate the association and predictive ability of tests performed at birth with adverse outcomes during the neonatal period and later life, with a view to guiding clinical practice regarding which tests should be performed, and how the results may be interpreted.

Objectives

The objectives of this thesis were as follows: a) to obtain summary estimates of association for neonatal tests and short and long term health outcomes. b) to obtain summary estimates of predictive ability of neonatal tests for short and long term health outcomes. c) to investigate the impact on the cost effectiveness of neonatal

hypothermia, to prevent cerebral palsy, of varying the umbilical cord pH threshold used to identify neonates in whom treatment should be given.

Methods

Systematic reviews were performed in accordance with current recommended methods. Decision-analytic modelling was used for the cost effectiveness analysis. For the systematic reviews, literature was identified from electronic searches, without language restrictions, studies were selected by 2 reviewers and data extracted to obtain 2 x 2 tables on the index test in question and mortality, or any measure of morbidity, throughout the life course. Methodological quality was assessed using pre-defined criteria from the STARD and QUADAS checklists. Meta-analysis was performed using the bivariate approach, to obtain summary measures of prognostic association (odds ratio) and predictive ability (sensitivity, specificity and likelihood ratios). A prediction interval was also calculated, which assesses the range in which the summary measure of a new study is likely to lie.

The decision-analytic model employed a decision- tree approach. Inputs to the model were systematic review accuracy data derived from this thesis, and effectiveness and cost data obtained from the literature. The perspective was that of the NHS with an 18 month time horizon. Deterministic and probabilistic sensitivity analyses were reported. The main outcome was cost per case of cerebral palsy avoided.

Results

Main findings of reviews of prognostic association and predictive ability

47338 citations were identified as potentially relevant for the included reviews. 218 papers and 26704980 individuals were finally included. The following tests were included: umbilical cord pH and base excess; any standard used to define low birth weight (including absolute birth weight at any reported threshold e.g. <2.5kg, population and customised centile charts, and ponderal index); and Apgar score.

51 papers were included in the review of umbilical cord pH and base excess, 92 in the birth weight standards review and 87 in the Apgar score review. The total number of individuals included per review was 479022 (cord pH), 3690080 (Apgar) and 23051541 (birth weight standards). Most studies were of a retrospective cohort design, and were of high or moderate quality according to predefined criteria. All of the neonatal tests assessed had a strong association with neonatal mortality, however even where the association was strong, the sensitivity and negative likelihood ratios were generally poor, indicating that a negative test does not change the odds of an adverse outcome. Most tests and thresholds examined also showed an association with neonatal morbidity and childhood cerebral palsy. However the magnitude of the association was smaller. It was only possible to assess the association between adult outcomes and low birth weight, and in this case there was no consistent relationship.

Findings of decision-analytic modelling

Decision-analytic modelling demonstrated that, when compared to a higher threshold of <7.10, the pH threshold of <7.00 used in current practice as one of the factors to indicate a neonate should receive therapeutic hypothermia is more effective and less costly and therefore the dominant strategy. Comparison with other thresholds was

not possible due to a lack of accuracy and effectiveness data. There were several limitations to the model in terms of data inputs and model structure, which restrict interpretation of the results of the cost-effectiveness of neonatal cooling overall.

Conclusions

The reviews in this thesis have shown that low umbilical cord pH, birth weight and a low Apgar score have significant associations with a variety of adverse outcomes and are therefore important tests to consider. However, given the fact that in most cases a negative test was not found to change the likelihood of an adverse outcome, and the inability to fully assess the performance of tests in population subgroups or compare multiple tests in the same population, there are many unanswered questions regarding their use and interpretation in clinical practice. The main recommendations of this thesis are thus for future research, to clarify these issues. Such research could take the form of individual patient data meta-analysis, or a large cohort or population registry study by which pertinent data on all tests and outcomes of interest could be examined, pH and birth weight as a continuous measure be assessed, and appropriate subgroups and potential confounding factors accounted for. Future studies should adhere strictly to the STARD checklist when reporting to facilitate meta-analysis. Decision–analytic modelling, also using individual patient data, could then be performed to assess the cost-effectiveness of tests in combination with available treatments or management strategies, to inform NHS practice.

DEDICATION PAGE

This thesis is dedicated to my husband Dave and sons Danny and Jamie.

ACKNOWLEDGEMENTS

This thesis could not have been completed and submitted without the help and support of my colleagues, not only at the University of Birmingham but also the University of Nottingham.

I would like to thank Dr Katie Morris for the many years of friendship, introducing me to the Academic department at Birmingham Womens' Hospital, becoming a co-supervisor for this thesis and the teaching, guidance and support that has enabled it to be completed.

Professor Khan, my supervisor, for his excellent ideas, expertise and opportunities to develop my skills during my period of study. Dr Morris and Professor Khan also were responsible for developing the grant application that secured funding for this work from the Mary Crosse fellowship.

Professor Tracy Roberts, my co-supervisor, for not only helping with the decision-analytic model based economic analysis and the excellent teaching provided within her department, but also encouragement and guidance with the thesis as a whole.

I thank Dr Richard Riley for the invaluable help with making statistics easy to understand and enjoyable, and for his expertise in prognostic research which has ensured that the analyses in this thesis were performed to the highest possible standard.

There were also colleagues not already mentioned who helped by being second reviewers on some of the systematic reviews: Dr Zainab Laftah, Dr Margreet Teune, and Dr Sofia Ahmad.

Friends and colleagues who helped with the translation of the papers: Miss Victoria Harold, Dr Amelia Hui, and Dr Petra Deering.

I confirm that for all chapters where a publication involved other authors I was involved at all stages for design, data collection, interpretation and writing of the papers. I was responsible for the design of the protocol for the individual reviews and performing the literature searches, obtaining the articles, performing literature selection, data extraction and quality assessment. I performed all of the statistical analyses with support from Dr Richard Riley. The interpretation of results was my own. For the economic evaluation, I identified the effectiveness data from the literature, performed the data collection around the costs of the test and intervention and the work to determine the cost of a non-encephalopathic infant. I built the tree and performed all of the analyses with the support of Professor Tracy Roberts. The interpretation of results was my own.

I would like to thank the Mary Crosse foundation for funding this work.

Finally I must thank my family, particularly my parents for their support and babysitting that has enabled this thesis to be completed, my grandmother for always taking an interest and who despite her best efforts did not see me finish it, and my husband Dave, for the sacrifices he has made for me and without whose love and understanding nothing would be possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABPM	Ambulatory Blood Pressure Monitoring
aEEG	Amplitude integrated Electroencephalogram
AGA	Appropriate for gestational age
AST	Aspartate transaminase
ALT	Alanine aminotransferase
β	Beta
BHR	Birth weight to head circumference ratio
BSID	Bayley scales of Infant Development
BMI	Body Mass Index
CDC	Centre for Disease Classification
CEAC	Cost Effectiveness Acceptability Curve
CI	Confidence Interval
CO_2	Carbon dioxide
CPAP	Continuous positive airway pressure
CR	Creatinine
CRIB	Clinical Risk Index for Babies

CS	Caesarean section
CTG	Cardiotocograph
DARE	Database of Reviews of Effectiveness
DFLY	Disability free life year
DNA	Deoxyribonucleic acid
ECG	Electrocardiogram
EPI	Estimated Prediction Interval
ERS	Examiners Rating Scale
FHR	Fetal heart rate
FHSA	Family Health Service Authority
FN	False negative
FP	False positive
g	Grams
GP	General Practitioner
H^+	Hydrogen ion
HCO_3^-	Bicarbonate
H_2CO_3	Carbonic acid
HIE	Hypoxic Ischaemic Encephalopathy

H_2O	Water
HRG	Healthcare Resource Group
ICD	International Centre for Disease
ICE	Infant Cooling Evaluation
ICER	Incremental Cost Effectiveness Ratio
IGF-1	Insulin-like growth factor 1
IGF-2	Insulin-like growth factor 2
IPD	Individual patient data
IQ	Intelligence quotient
IUGR	Intrauterine growth restriction
IVH	Intra-ventricular haemorrhage
kg	Kilograms
Ln	Log
LR	Likelihood Ratio
m	Metres
MeSH	Medical Subject Headings
mg	Milligrams
mg/dL	Milligrams per decilitre

mEq/L	Milliequivalent per litre
mmHg	Millimetres of mercury
mmol/L	Millimoles per litre
MRC	Medical Research Council
NEC	Necrotising Enterocolitis
NHS	National Health Service
NIDDM	Non-Insulin Dependent Diabetes Mellitus
NNU	Neonatal Unit
NPV	Negative Predictive Value
OGTT	Oral Glucose Tolerance Test
ONS	Office of National Statistics
OR	Odds ratio
PD	Prism Diopters
PICOS	Population, intervention/ index test, comparator, outcome, study design
PPV	Positive Predictive Value
PSA	Probabilistic Sensitivity Analysis
PVL	Periventricular leukomalacia
QUADAS	Quality Assessment of Diagnostic Accuracy Studies

RCT	Randomised Controlled Trial
RDS	Respiratory distress syndrome
RR	Relative risk
SD	Standard deviation
SE	Standard Error
SCPE	Surveillance of Cerebral Palsy in Europe
SGA	Small for gestational age
SNAPPE	Score for Neonatal Acute Physiology- Perinatal Extension
STARD	Standards of Reporting for Diagnostic accuracy studies
TN	True negative
TP	True positive
UK	United Kingdom
U/L	Units per litre
USS	Ultrasound scan
WRAT	Wide range achievement test
YCI	Yale Childrens Inventory

PUBLICATIONS FROM THIS THESIS

Malin GL, Morris RK, and Khan KS. Mary Crosse project: Systematic reviews and grading the value of neonatal tests in predicting long term outcomes. *BMC Pregnancy and Childbirth 2009; 9:49.*

Malin GL, Morris RK, Khan KS. Umbilical cord pH in the prediction of neonatal and long term morbidity: a systematic review of the literature and meta-analysis*. Archives of Disease in Childhood: Fetal and Neonatal Edition 2009; 94 (supplement 1): Fa 26*

Malin GL, Morris RK, Khan KS. What is the relationship between umbilical cord pH and perinatal and long term outcome? A systematic review and meta-analysis to evaluate strength of association. *BMJ 2010; 340: c1471.*

Malin GL, Morris RK, Riley RD, Teune M, Khan KS. Should we forget about centile charts? Comparing definitions of fetal growth restriction to predict adverse outcome. *Archives of Disease in Childhood: Fetal and Neonatal Edition 2012; Suppl 1. A111.*

CHAPTER 1: INTRODUCTION

1.1 Aims and objectives of this thesis

The aim of this thesis was to examine the prognostic association and predictive ability of neonatal tests for short and long term health outcomes.

The objectives were as follows:

1. To obtain summary estimates of the prognostic association of neonatal tests with short and long term health outcomes.
2. To obtain summary estimates of the predictive ability of neonatal tests for short and long term health outcomes.
3. To investigate the impact on the cost effectiveness of neonatal therapeutic hypothermia, to prevent cerebral palsy, of varying the umbilical cord pH threshold used to identify neonates in whom treatment should be given.

1.2 Outline of this thesis

This thesis has been divided into two parts:

PART A: Systematic reviews of association and predictive ability of neonatal tests for short and long term outcomes.

PART B: Summary of existing literature for neonatal hypothermia to treat hypoxia, and decision-analytic modelling

Appendices and References

1.3 Methods

This thesis employed systematic reviews and meta-analyses to calculate the prognostic association and predictive ability of neonatal tests for a variety of short and long term health outcomes. Focusing on the umbilical cord pH, which is a measure of neonatal hypoxia, available data regarding neonatal cooling to prevent neurodevelopmental problems in these infants was combined with the systematic review prediction data in a decision-analytic model, to explore the cost-effectiveness of varying the threshold of the test at which treatment is given.

CHAPTER 2: BACKGROUND

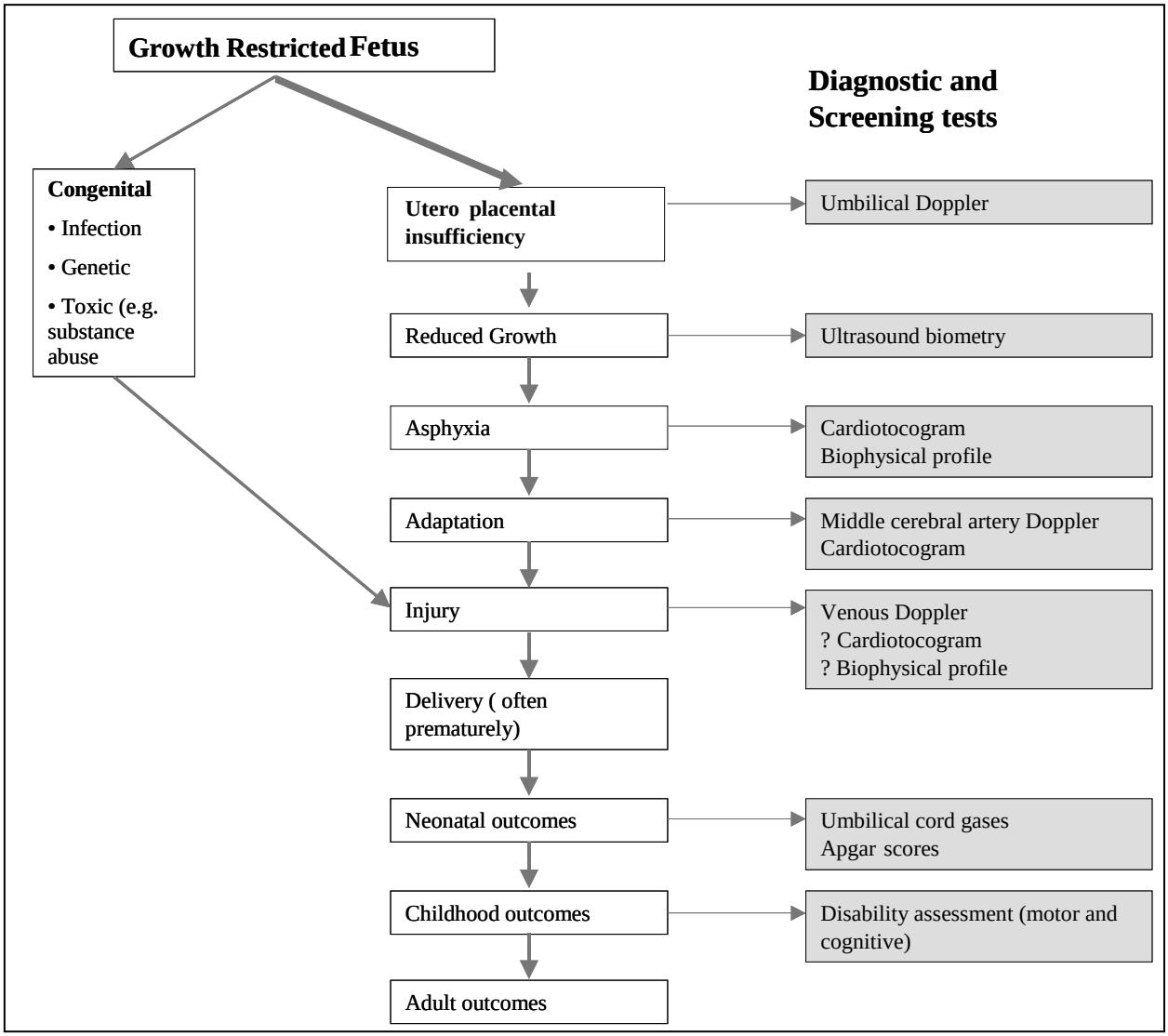
2.1 The purpose of neonatal testing and scope of the problem

There is a growing interest in health across the life course. 'Living a long and healthy life' has been identified as a priority research theme by the Medical Research Council (MRC),[1] highlighting the importance of identifying individuals at risk of adverse health outcomes, facilitating intervention. Events in utero have been linked with a variety of diseases throughout life, from the neonatal period to adulthood. Such conditions may be severe and life limiting. However, there is a lack of consensus regarding the ability of neonatal tests to predict these outcomes.

Cerebral palsy, for example, is often linked to adverse events occurring in the antenatal or intrapartum period. This is a permanent condition, arising from an insult to the developing brain, resulting in non-progressive abnormalities of movement and posture. It varies in severity, but an individual with this diagnosis is likely to have lifelong needs in terms of health and social care, with the associated costs to the health service, and a significant impact on quality of life. The prevalence of cerebral palsy within the UK is estimated at 2 to 2.5 per 1000 live births.[2] There were 723165 live births in England and Wales in 2010,[3] which would mean a total of 1446 to 1808 individuals affected with cerebral palsy in that year alone.

Tests performed immediately following delivery of a baby not only summarise the condition at that moment, but also reflect events in utero and during labour, which have implications for obstetric practice. Figure 2.1 illustrates the pathway by which neonatal tests are related to antenatal and postnatal events.

Figure 2.1 The pathway by which neonatal tests are related to antenatal and postnatal events



This thesis aims to evaluate the association and predictive ability of tests performed at birth for adverse outcomes during the neonatal period and later life, with a view to guiding clinical practice regarding which tests should be performed, and how the results may be interpreted.

2.2 The relationship between fetal and neonatal wellbeing

2.2.1 Oxygen transfer and metabolism

Fetal oxygen levels are maintained through diffusion of oxygen from the maternal circulation via the placental interface. Fetal haemoglobin has a higher affinity for oxygen than adult haemoglobin, and this allows saturation with oxygen at relatively low oxygen tensions. Under normal circumstances, fetal blood is slightly hypercarbic and acidotic in comparison to maternal blood.[4] Carbon dioxide (CO_2) is related to the concentration of hydrogen ions, and therefore acidity of the extracellular fluid within the body, as follows:[5]

$$CO_2 + H_2O \rightleftharpoons H_2CO_3 \rightleftharpoons H^+ + HCO_3^-$$

Bicarbonate (HCO_3^-) acts as a buffer to maintain the pH balance of the blood. The base excess or deficit describes the quantity of bicarbonate within the plasma. Under normal circumstances, fetal arterial deoxygenated blood courses through placental exchange vessels and gives up hydrogen ions (H^+) and CO_2 to the maternal circulation and acquires oxygen, which returns from the placenta to the fetal circulation via the umbilical vein.[4]

Any disturbances to this process will result in fetal hypoxaemia, defined as low oxygen levels in the fetal blood, and hypercarbia, with a subsequently reduced pH.

However, in the early stages of a hypoxic event no metabolic changes occur at the tissue level. This would give a low pH in the presence of a normal base excess, a so called ‘respiratory’ acidosis. The most common causes of this are a sudden decrease in placental or umbilical perfusion. This may result from umbilical cord compression or uterine hyperstimulation. Maternal causes include hypotension, for example following regional anaesthesia, and also maternal hypoxia or hypoventilation, such as in asthma or massive pulmonary embolus.[6] Reversal of the underlying cause and a restoration of circulation will lead to normalisation of the acid-base balance.

If the episode of hypoxia is prolonged, the fetal tissues switch to anaerobic metabolism, resulting in the accumulation of lactic acid. This leads to a rise in lactate levels, a loss of bicarbonate and therefore a larger base deficit, a ‘metabolic’ acidosis.[7] This situation may arise from a persistence of an acute event, as described above, or chronic uteroplacental hypoperfusion. It has been suggested that adverse health outcomes are associated with metabolic, rather than respiratory acidosis.[8;9]

Should hypoxaemia occur, the fetal cardiovascular system will respond to the insult through transient bradycardia, an increase in arterial blood pressure, and a redistribution of the combined fetal ventricular output in favour of the adrenal glands, heart and brain, at the expense of perfusion to the peripheral circulation.[10]

Consequences of hypoxia

At a tissue level, hypoxic ischaemia initiates energy depletion, the accumulation of extracellular glutamate, and activation of receptors, leading to a deleterious cascade of events resulting in cell death.[11] The effects may be multisystem, resulting in renal injury, myocardial dysfunction, and encephalopathy. Death may result if severe,

irreversible tissue damage occurs. Peripartum asphyxia affects 3 to 5 per 1000 live births in developed countries, with moderate or severe hypoxic ischaemic encephalopathy occurring in 0.5-1 per 1000.[12] The long term consequences include cerebral palsy, neurodevelopmental delay and learning difficulties.[13-15] However, these outcomes are unpredictable due to the fact that different areas of the developing brain are susceptible to damage at different times.[16] In infants born at term (≥ 37 weeks gestation), neuronal injury predominates, and the hippocampus, deeper layers of the cerebral cortex and cerebellar Purkinje cells are the most frequently injured.[17] In premature infants, the cerebral white matter is the major site of injury, resulting in periventricular leukomalacia (PVL), but primary or secondary injury to cortical or deep grey matter may also occur.[17] However, the clinical consequences are unpredictable, resulting from the variation in site and severity of lesions and the phenomenon of 'plasticity', resulting from an ability of the brain to functionally adapt.[18]

2.2.2 Labour and delivery events

Uterine contractions transiently reduce the uterine blood flow and impair gaseous exchange. Under normal circumstances, the fetus has a period of recovery in between contractions, where circulation is restored, allowing a normal acid-base balance to be maintained. A small drop in pH level over the course of labour and delivery is expected,[19] and the following mean umbilical cord values and standard deviations have been quoted as normal post-delivery of a term neonate:[20]

Arterial pH 7.27+/- 0.07; Base excess (mEq/L) -2.70+/-2.8

Venous pH 7.34 +/- 0.06; Base excess (mEq/L) -2.40+/-2.0

Immediately following delivery, under normal circumstances the neonate will begin breathing within a minute of birth, due partly to the drop in pH occurring due to labour, and also the stimulation of neuroreceptors in the suddenly cooled skin.[21] This response occurs at the brainstem level. Failure to initiate breathing spontaneously may result from maternal anaesthesia or opioids, which cause depression of the fetal respiratory centre, however asphyxia may also cause apnoea. Animal models have demonstrated that following interruption of blood supply to the umbilical cord, a cessation of breathing movement occurs (primary apnoea), followed by a period of gasping, during which respiratory effort can be restored by sensory stimuli. If the hypoxia persists, a secondary phase of apnoea is entered, followed by eventual cardiac arrest.[22]

Depression of the neonatal heart rate immediately after delivery is most often caused by hypoxia; other causes include cardiac arrhythmia, structural developmental anomalies and infection.[23] If hypoxia is the cause, re-oxygenation stimulates an increase in heart rate, unless there has been significant myocardial damage as a result of severe or prolonged hypoxaemia.[23] Neonatal colour is dependent on both circulation and heart rate: compromise of either of these may result in cyanosis or pallor.[22] Neonatal muscle tone following delivery is reliant on higher cerebral function. Passive flexor tone appears between 28 and 34 weeks gestation and matures from the legs upwards, therefore normal term babies will lie with their limbs flexed and adducted.[24] Normal spontaneous movements of the limbs will be varied, and babies should open their hands and move their fingers. Normal reflexes that are present, controlled at brainstem level, include the Moro reflex, sucking, rooting and stepping.[24]

Hypotonia at birth may result from hypoxic ischaemic encephalopathy, other ischaemic or haemorrhagic brain lesions, and in association with chromosomal abnormalities or metabolic disorders. Peripheral hypotonia may be caused by specific myopathy or neuropathy.[25] Drugs administered to the mother may cause transient neonatal hypotonia.[26]

2.2.3. Intrauterine fetal growth

Normal growth is characterised by periods of tissue and organ growth, differentiation and maturation. The process is dependent on maternal provision of substrates (including amino acids, glucose, fats and oxygen), adequate placental transfer, and genetic determination of fetal growth potential. In the first 16 weeks of in utero life, growth occurs predominantly by hyperplasia and increasing cell numbers, followed between 16 and 32 weeks by both cellular hypertrophy and hyperplasia; after 32 weeks gestation hypertrophy predominates.[27] During the first 16 weeks of fetal life, there is little variation in growth rate, and this is largely under genomic control.[28] There is a consensus that insulin like growth factor 2 (IGF-2) is an important paracrine factor in embryonic growth.[28] In later gestation, IGF-1 is the predominant growth regulator, produced by the fetal liver and other tissues. The main factor stimulating production of IGF-1 is fetal insulin, which in turn is regulated by fetal glucose availability.[29] The growth velocity in later gestation is much more variable. There is a genetic influence, predominantly relating to the genes determining maternal habitus, with paternal influence being less significant. However the dominant factor is the supply of nutrients and oxygen to the fetus, which is determined by a balance of maternal health and homeostasis, uteroplacental circulation, placental transfer and metabolism, and umbilical blood flow.[28]

Normal birth weight for a baby born at term (≥ 37 weeks gestation) is considered to be 2500 to 4000 grams (g). However, this range should be interpreted with caution, as certain babies will be genetically programmed to be smaller and therefore will fall outside this range while still achieving their growth potential. Equally, a baby who is destined to be larger may undergo a pathological process of growth restriction in utero, yet still fall into the normal range at delivery.

The causes of intra-uterine growth restriction

The incidence of intrauterine growth restriction varies according to the definition used, but is estimated to be 5-7%[30] This represents a heterogeneous group of aetiologies. The normal growth process may be disturbed by any factors disrupting maternal nutrient provision, placental transfer or genetic programming. These are summarised in table 2.1:

Table 2.1 Aetiology of intrauterine growth restriction

Maternal factors	**Fetal factors**	**Placental**
Pre-eclampsia/ hypertension	Multiple pregnancy	Insufficiency
Severe chronic disease (e.g. cardiac/renal/respiratory)	Infection	Abruption/ infarction
Anaemia	Chromosomal disorders (e.g. trisomy 21, 18)	Vascular anomalies
Smoking	Infection (e.g. toxoplasmosis, rubella)	Infection
Infection	Congenital anomaly	
Substance abuse		
Low body mass index/ malnutrition		

Placental insufficiency may occur in association with maternal pre-eclampsia, and results from failure of adequate trophoblast invasion into the maternal spiral arteries, limiting the potential for nutrient transfer.[31] This is thought to occur in approximately three percent of pregnancies.[32] Maternal ill-health may alter the potential for nutrient or oxygen provision within the circulation. This, along with placental insufficiency, will extrinsically cause the fetus to fail to reach its growth potential. Chromosomal disorders, congenital anomalies and infections either lead to altered growth potential within the fetus, or intrinsic failure to achieve this. The latter tend to cause a pattern of symmetrical growth restriction, where both fetal head and abdominal circumference are small, and this may be begin in the first or second trimester, resulting in decreased cell number and/ or size.[30] Placental insufficiency or maternal vascular factors tend to cause asymmetric growth restriction, whereby the fetus has a smaller abdomen compared to head circumference. This process tends to become apparent in the third trimester, and results from impaired cellular hypertrophy. The growth asymmetry results from adaptation in the fetus to protect circulation to the brain and heart therefore diverting blood away from the peripheries and abdominal organs.[30]

Consequences of intra-uterine growth restriction

In severe cases, intra-uterine growth restriction may result in stillbirth or neonatal death. [33] In survivors, it has been associated with neonatal morbidity, resulting from oxygen and substrate deprivation during intrauterine life, such as hypoxic ischaemic encephalopathy, polycythaemia, hypothermia and hypoglycaemia.[34] They may also be at risk of later life consequences such as cerebral palsy and neurodevelopmental delay.[35;36]

In addition, intrauterine growth restriction has been linked with a number of adverse health outcomes in adulthood, including cardiovascular disease, diabetes mellitus and hypertension.[37;38] The 'fetal origins hypothesis' was first proposed in 1986 when Barker et al demonstrated an inverse relationship between birth weight and adult cardiovascular disease.[39] The theory suggesting that malnourishment in utero results in fetal programming, whereby biological pathways are altered, resulting in increased susceptibility to disease in adult life. A number of mechanisms have been proposed, including metabolic effects, such as disruption of lipid metabolism in the adipose tissue and liver, and β- cell dysfunction leading to altered glucose homeostasis, resulting from fetal adaptations to malnutrition.[40] Non-metabolic effects such as increased apoptosis, leading to altered organ morphology, have also been proposed.[40] Epigenetic programming, whereby changes in gene transcription occur as a result of mechanisms other than changes to the underlying DNA sequence, have also been linked to this hypothesis, with suggestions that intrauterine growth restriction may alter chromatin associated with important genes, varying their expression.[40;41]

Since the 'fetal origins' or 'Barker hypothesis' was first described, numerous studies have demonstrated an association between low birth weight and morbidity and mortality from a variety of conditions in childhood and adult life.[37;42-48] However, the results have not always been consistent.[38;49-51] The initial evidence for the Barker hypothesis has been criticised for failing to account for important potential confounders such as gestational age and socio-economic class, and as such is not universally accepted.[52] Indeed, should the association be genuine, it may not follow that it is also causative. The 'fetal insulin hypothesis' suggests that the relationship

between low birth weight and non-insulin dependent diabetes mellitus (NIDDM) may be attributable to common genetic factors relating to birth weight and NIDDM risk, rather than the diabetes as a consequence of intrauterine growth restriction.[53]

2.2.4 Prematurity

Premature infants (born at <37 weeks gestation) may differ significantly from term infants, depending on the gestation at delivery. Due to both low birth weight and immaturity of organ systems, there are a number of physiological differences from term infants, and they are at higher risk of a number of pathological processes; the combination of these features leads to differences in the clinical picture at delivery and increased risk of adverse outcomes.

Neurologically, primitive reflexes such as the Moro response and asymmetric tonic neck reflex do not fully appear until 33 to 35 weeks gestation, and rooting until 28 weeks. Premature babies tend to adopt an extended posture at rest.[24] They are at higher risk of intraventricular haemorrhage (IVH), periventricular white matter injury and hydrocephalus.[54]

Due to immaturity in central and peripheral chemoreceptors, control of respiration is less well developed, and premature infants will frequently display pauses in respiration. If they are accompanied by bradycardia, cyanosis or pallor, or last at least 20 seconds, this is considered significant. In the majority of infants they improve as the infant matures, however they may be related to an underlying neuropathologic process such as hypoxic ischaemic encephalopathy (HIE) or IVH.[55] As a consequence of delay in surfactant production within the immature lungs, premature infants are also at higher risk of respiratory distress syndrome.[54]

Other complications arising from preterm delivery include necrotising enterocolitis, increased susceptibility to infection, patent ductus arteriosus, fluid and electrolyte imbalance, retinopathy of prematurity and hypothermia and hypoglycaemia.[54] The risks of long term problems including neurodevelopmental delay, cerebral palsy, bronchopulmonary dysplasia, hypertension and increased insulin resistance are also increased.[54]

2.3. Available neonatal tests

2.3.1 Summary of neonatal tests available in the UK

The tests, and the context in which they are usually performed, are summarised in Table 2.2.

Table 2.2 Available tests of neonatal wellbeing according to current UK practice

Test performed	Routine	'High risk' infants only	Research tool	Varies according to local practice
***Measure of fetal growth*:**				
Absolute birth weight	√			
Birth weight according to population centile chart	√			
Birth weight according to customised centile chart				√
Ponderal Index (birth weight (kg)/ length (m)^3				√
Fetal growth ratio (observed weight/ expected for population)			√	
Head circumference (cm)	√			
Skinfold thickness		√		
Weight for length z score			√	
Mid-arm circumference/ head circumference ratio			√	
Measurement of fetal hypoxia:				
Umbilical artery cord pH		√		
Umbilical vein cord pH		√		
Umbilical artery base excess		√		
Umbilical vein base excess		√		
Arterial pH within 1 hour of birth		√		
Serum lactate		√		
Overall measure of wellbeing:				
Apgar score	√			
Admission to Neonatal Unit required?	√			
CRIB (infants admitted to NNU)				√
SNAPPE score (infants admitted to NNU)				√

2.3.2 Current service provision: a description of the context in which the tests are performed and the results utilised

Indicators of fetal hypoxia

Blood sampling from the umbilical cord at delivery, for measurement of pH and base excess, gives an indication of whether the neonate has become hypoxic prior to delivery. Current UK guidelines for antenatal care recommend that this is not done routinely for all deliveries, rather that it should be reserved for infants who are suspected to be at higher risk of compromise, including those who have undergone an instrumental delivery, or had abnormalities on antenatal fetal heart rate monitoring, or any other cause for concern.[56] However, in some units in the UK, umbilical cord pH testing is performed routinely at delivery.[57]

The way in which these results are utilised clinically differs widely. In some units, a low pH in an otherwise well baby would prompt admission to the neonatal unit. In others, no further action would be taken. The most common course of action is to take the pH into consideration along with other tests, or clinical evidence of overall neonatal wellbeing,[57] to determine the need for further monitoring or treatment including neonatal therapeutic hypothermia.[58]

Other tests to determine the acid-base status of a neonate include arterial blood sampling within the early neonatal period, and measurement of the serum lactate level. These would generally be reserved for babies who required admission to the neonatal unit, but the results would be used to guide further monitoring or treatment, in the same way as the umbilical cord pH result.

Birth weight

At present, all infants are weighed routinely and the head circumference is measured at birth. Depending on local practice, the length may also be measured. The weight for gestational age is then plotted on a centile chart, which is based on a range around normal values for the general population.[59] Depending on local practice, the ponderal index (birth weight (kg)/ length (m^3)) may be calculated, or a customised centile chart (normal range defined according to the mother's height, weight, parity and fetal sex) used.[60] If an infant is deemed to be of low birth weight, for example less than 2.5kg at a gestation of 37 weeks or greater, this may be used to indicate the need for further intervention and monitoring such as blood sugar testing and admission to the neonatal unit.

Overall measures of wellbeing

The Apgar score is routinely recorded for all infants at delivery by the attending midwife or doctor. Neonatal characteristics are evaluated as described in Table 2.3, including heart rate, respiratory effort, reflex responsiveness, muscle tone, and colour, and a total score assigned, in a range from zero to ten. The score is usually recorded at one, five and ten minutes following delivery.[61] The initial purpose of the score, when first introduced by Virginia Apgar in 1953, was to focus attention on the newborn infant and identify those in need of resuscitation.[62] The Apgar score may be reduced by any factor that causes compromise to the neonate, for example prematurity, hypoxia, and maternal opioids prior to delivery.[63] The change in Apgar score over time documents neonatal recovery and response to resuscitation. Again, the action taken for a low Apgar score at delivery, besides immediate resuscitation,

varies according to local practice and how long from delivery the low score persists for. Some would admit all babies with a ten minute Apgar score less than seven to the neonatal unit, whereas the majority would perform an additional review of the baby at a later time, and a minority would take no further action. [57]

Table 2.3 The Apgar score [64]

Sign	Score 0	Score 1	Score 2
Heart rate	Absent	<100/min	>100/min
Respiration	Absent	Weak	Good cry
Muscle tone	Flaccid	Some flexion	Well flexed
Reflex response (to skin stimulation of skin of feet)	No response	Grimace	Cry or active withdrawal
Colour	Pale/ blue	Blue extremities	Pink

Other measures that may be considered an overall indicator of neonatal wellbeing include whether the baby is admitted to the neonatal unit or not. On admission, an additional scoring system such as the Clinical Risk Index for Babies (CRIB), or SNAPPE score, may be applied to assign a risk of adverse outcome depending on a combination of factors such as temperature, pH, blood pressure or base excess.[65;66] Use of these scores varies according to local practice.

2.3.3 Selection of tests to include in this thesis

Three tests were selected from those available to be the focus of systematic reviews performed in this thesis. Umbilical cord pH or base excess, birth weight and Apgar

score are tests that are performed immediately following delivery, and are not only used to inform the future care of the neonate, but also as bench marks for obstetric care, both within clinical practice and frequently as outcome measures in obstetric clinical trials.[67;68 69;70] In this context they are considered a surrogate for neonatal and long term morbidity, however due to the uncertainty surrounding this assumption observed effects are often viewed with suspicion.

2.3.4 Summary of the current evidence for the association of neonatal tests with short and long term outcomes

Umbilical cord pH

The occurrence of perinatal asphyxia can be measured by the presence of fetal acidosis, determined by umbilical cord pH at birth.[7] Cerebral palsy is thought to occur more frequently at an arterial cord pH of <7.00 and a base deficit ≥ 12mmol/l.[71] However, such criteria have been derived through consensus, not through evaluation of collated evidence summaries in the field,[71] leading to some uncertainty about their use.[72] This is because existing studies of the association between pH levels and outcomes have drawn inconsistent inferences, in part due to the different parameters measured, including arterial or venous pH and base excess, the variety of outcomes evaluated and the different thresholds used to define abnormality.[73-76]

Previous reviewers have examined the relationship between fetal acidosis and the outcomes neonatal death and cerebral palsy, and found a significant association for both.[77] However, this review was performed a decade ago, during which time new studies on the subject have been published and guidelines produced regarding the methodology and reporting of systematic reviews, including quality assessment of

included studies, which were not in widespread use at the time of its publication.[78] Substantial uncertainty remains about the value clinicians may attach to acidosis in the clinical management of neonates, and the long term implications of the umbilical cord pH result.

Birth weight standards

As described in Table 2.2, there are a variety of different definitions of low birth weight, which attempt to capture cases of intrauterine growth restriction, and identify infants most at risk of adverse outcome. These include population based centile charts, the most commonly used threshold being the tenth centile;[21] customised charts where the mother's BMI and ethnicity are used to calculate individualised growth centiles;[22] and ponderal index which takes into account the neonatal weight and length.[23] The published associations between each standard for defining growth restriction and adverse outcome vary, and there is no current consensus regarding the best method.[24]

A number of systematic reviews have been performed to assess the relationship between low birth weight and a variety of health outcomes in childhood and adult life. However, the definitions of low birth weight have varied and results have been mixed, with pre-term infants included in the analysis, which may confound the results. [37;79;80] No existing systematic reviews comparing definitions of low birth weight have been identified.

Apgar score

A group of neonates with a low Apgar score defined by a conventional cut-off will include neonates with asphyxia, but also a low score due to maternal medication,

delivery method or anaesthesia, and therefore the interpretation of these results has been questioned.[81] Also, there is a lack of consensus regarding the threshold that defines a 'low' score, and the timing of measurement to which the greatest importance can be attached.[82] A systematic review of the association of Apgar score with adverse outcome demonstrated a significant association between a low score and both neonatal mortality and later cerebral palsy.[77] However, this review was published a decade ago, and does not differentiate between pre-term and term born infants, between whom the Apgar score and adverse outcomes may vary widely.[83]

A comprehensive systematic review of the literature surrounding these tests will clarify the existence and strength of a relationship between an abnormal test and adverse outcomes throughout the life course, and guide clinical practice regarding the use of different tests and thresholds to predict these.

2.4 Development of this thesis

The Mary Crosse Fellowship (Birmingham Women's Hospital NHS trust) funded an evidence synthesis project to examine the association between neonatal tests and health outcomes. The author worked on this project performing systematic reviews and health economic evaluation, and also received formal training at the University of Birmingham in Health Technology Assessment, Health Economic Evaluation, Advanced Statistics and Decision Modelling; and the Oxford Centre for Evidence Based Medicine in Evidence Based Diagnostics.

CHAPTER 3: RESEARCH QUESTIONS ADDRESSED IN THIS THESIS

3.1 Research questions addressed in this thesis by systematic review

1. What is the association of umbilical cord pH or base excess with neonatal and long term adverse outcomes?
2. What is the predictive ability of umbilical cord pH or base excess for neonatal and long term adverse outcomes?
3. What is the association of different standards for defining low birth weight with neonatal and long term adverse outcomes?
4. What is the predictive ability of low birth weight standards for adverse outcome?
5. What is the association of the Apgar score at birth with neonatal and adverse outcomes?
6. What is the predictive ability of the Apgar score for adverse outcomes?

3.2 Research questions addressed in this thesis by decision- analytic modelling

1. Which threshold of umbilical cord pH is the most cost-effective in combination with treatment with neonatal hypothermia to reduce the risk of cerebral palsy following intrapartum hypoxia?

PART A: SYSTEMATIC REVIEWS OF THE ASSOCIATION AND PREDICTIVE ABILITY OF NEONATAL TESTS FOR SHORT AND LONG TERM ADVERSE OUTCOMES

CHAPTER 4: METHODS FOR SYSTEMATIC REVIEWS OF ASSOCIATION AND PREDICTIVE ABILITY

4.1 Introduction

The systematic reviews performed in this thesis evaluate the association and predictive ability of tests performed at birth for neonatal and long term adverse outcomes. The neonates on whom the tests were performed were from different population spectrums and, where possible, pre-term (<37 weeks gestation) and term born (≥ 37 weeks) were analysed separately to reflect the different risk of adverse outcomes between these populations. The systematic reviews used a common methodology which is detailed in this chapter. Where specific methodology differed, this is given in the chapter relevant to each test.

All systematic reviews were performed using a prospective protocol and widely recommended methods.[84-88] The following steps were taken in all cases:

i) Framing the question

ii) Study identification

iii) Quality assessment

iv) Data extraction

v) Analysis and interpretation of data

This work has been published by Malin GL, Morris RK, and Khan KS. Mary Crosse project: Systematic reviews and grading the value of neonatal tests in predicting long term outcomes. *BMC Pregnancy and Childbirth 2009; 9:49.*

4.2 Framing the question

Each review should set a clear question in order to ensure that the objectives of the review are achieved. The question can be framed according to the 'PICOS' criteria, ensuring that all elements of the review question are clearly defined.[84] This includes the population, index test (or intervention), comparator (not used in these reviews), outcome measure and study design of interest; a well-structured question will determine components of the search strategy and therefore aid identification of relevant studies for inclusion. The common components of the question used in each of the reviews in this thesis are given below.

1. **Population** Live-born infants who have had the index test of interest performed at birth.

2. **Index test** Any measure of weight or growth at birth including: absolute birth weight (thresholds <2.5kg, <2.0kg, <1.5 kg); population or customised centile charts (thresholds < 10^{th} centile, < 5^{th} centile, <3^{rd} centile); ponderal index or other growth ratios; Umbilical cord blood examination for arterial or venous pH or base excess; Apgar score at 1, 5 or 10 minutes of age.

3. **Outcome** Any measure of compromise of neonatal, childhood or adult wellbeing such as: mortality, neonatal morbidity including hypoxic ischemic encephalopathy; childhood or adult motor disability; childhood or adult disease including diabetes mellitus, cardiovascular disease and hypertension.

4. **Study design** Randomised controlled trials or observational studies (including cohort and case control design) that allowed generation of a 2x2 table (true positives, false positives, false negative and true negatives) to compute an

estimate of the association and predictive ability between test result and outcomes. Studies with ≤ 5 individuals in total were excluded on account of unreliability.

4.2.1 Population

In all cases the population for inclusion was live born infants. In the birth weight standards review, this was further limited to infants born at term (≥ 37 weeks gestation) to avoid the confounding effect of prematurity on this group of tests. No other restrictions in terms of study setting or other risk factors for adverse outcome were applied, although these were considered in subgroup analyses where possible.

4.2.2 Index tests

The reasons for selection of the included tests are described in Section 2.3.3. Scoping searches were performed to confirm the availability of primary evidence for these tests and the outcomes of interest. To ensure that the assessment of each test was as comprehensive as possible all methods and thresholds described by authors in primary studies were included.

4.2.3 Outcome measures

In order to optimise the assessment of the tests, the outcome measures selected for inclusion were not limited to specific conditions *a priori*, rather the search strategies used a combination of specific adverse outcomes (including neonatal, childhood and adult mortality, learning difficulties, cerebral palsy, hypertension, diabetes mellitus) and open terms (e.g. morbidity) and any papers with the appropriate index test and

population and any measure of adverse outcome throughout the life course were included.

4.2.4 Study design

Randomised controlled trials (RCT) or observational studies (including prospective and retrospective cohort and case control design) that allowed generation of a 2x2 table (true positives, false positives, false negative and true negatives) to compute an estimate of the association and predictive ability between test result and outcomes. Studies with ≤ 5 individuals in total were excluded on account of unreliability. Case-control studies, whereby individuals are selected for inclusion on the basis of their disease status, are thought to be prone to spectrum bias.[89] This was accounted for in the quality assessment of included studies.

4.3 Study identification

The search strategies for each study were designed to capture primary studies with data regarding the association or predictive ability of each neonatal test with adverse outcomes throughout the life course, using elements of the question outlined in section 4.2.

The initial scoping search, outlined in Appendix 1, was performed by Dr R K Morris (see acknowledgments) in preparation for the application for fellowship funding for this project. The search was run in 2006 in three electronic databases (Medline, EMBASE, Cinahl) at which time 3698 potentially relevant citations were retrieved. The final searches, outlined in chapters 5 to 7, were performed by the author following appointment to the fellowship and decision of the final tests for inclusion in this thesis.

In all cases, a combination of MeSH headings, keywords and word variants for the index test in question were combined using the Boolean operator ‘OR’ for capturing citations of the relevant text. These were combined using ‘AND’ with a combination of MeSH headings, keywords and word variants to capture relevant outcomes. The search was restricted to human studies, but no language or methodological restrictions were applied. Pilot searches were performed in Medline before the final search strategy was applied to maximise capture of relevant studies.

Searches were performed from database inception until the search date. The following sources were used to search for relevant literature:

1. Medline and EMBASE (general bibliographic databases)
2. Specialised electronic databases: The Cochrane Library (including systematic reviews and clinical trials databases), DARE (Database of reviews of effectiveness), Medion, British Nursing Index (EBSCO)
3. Sources of ‘grey’ literature: Web of Science, OpenGrey, contact with experts in the field.
4. Hand searching of specialist journals
5. Reference checking of review articles and included studies
6. Contact with authors for clarification where 2 x 2 tables could not be obtained from published studies.
7. SCISEARCH and Web of Science to identify frequently cited articles and conference proceedings

A comprehensive database collating all citations was constructed using Reference Manager 11.0/12.0. The list of citations was initially scrutinised by the author, and full

articles of all citations that were felt on examination of title and abstract to be likely to meet the predefined selection criteria were obtained. Translations of articles in languages other than English were obtained (see acknowledgements). Final inclusion or exclusion decisions were made through their examination by the author, depending on their adherence to a pre-defined checklist based on population, index test, outcome measure, study design and ability to obtain data to populate 2 x 2 tables. The study selection and data extraction process was also repeated, in a random sample of 10% of the papers, by a second reviewer (see acknowledgments). Where any disagreements occurred, these were resolved by consensus or the input of a third reviewer (Professor K S Khan). A random sample rather than dual review of all papers was performed for pragmatic reasons, due to the large number of articles reviewed at each stage of the process. There was a high level of inter-reviewer agreement (98%) which was felt to justify deviating from the recommendation of reviewing in duplicate throughout.

All manuscripts were carefully examined to identify duplications in population. Where this was identified, the most recent and complete versions were selected.

4.4 Quality assessment

All articles meeting the selection criteria were assessed for methodological quality, defined as the confidence that the study design, conduct and analysis minimised bias in estimation of the association. Quality was assessed (by the author and in duplicate by a second reviewer where possible) using the complete Standards for Reporting of Diagnostic Accuracy (STARD) (Appendix 2) and Quality Assessment of Diagnostic Accuracy (QUADAS) (Appendix 3) checklists, which are validated for the reporting

and methodological quality of diagnostic test accuracy studies. The quality elements from these checklists which were felt to be most relevant for these reviews on prognostic tests, associations and predictive ability were selected.[90;91] A quality score was not assigned as this been shown to give flawed results.[92] Cohort study design was considered to be superior to case- control. A study was rated high quality if it had at least four of the following items: adequate description of population; adequate description of the index test and outcome measure; consecutive recruitment; prospective recruitment; > 90% completions of follow up; appropriate outcome measurement; blinding of the investigators performing the outcome measure and a statement regarding the use of intervention between the index test and outcome. A study was deemed to be of medium quality when only three criteria were met and low if two or less were adhered to. Description of the population was considered adequate if the gestational age at delivery and the setting from which the population was recruited were reported, and information regarding risk factors or co-morbidities for adverse outcome given. An 'unclear' grading was given if some of this information was missing, for example the gestational age of the population was reported but not the study setting. Consecutive or random recruitment were considered ideal, and a prospective study design was thought to introduce less bias than retrospective methods. The index test was considered to be described in full if the timing, method and personnel taking the test were specified. Adequate description of the outcome measure required specification of the exact nature of the condition in question, the threshold used to determine whether an individual had the condition of interest or not, and who performed the test to determine disease status. Blinding of that individual to the index test result, and a description of any treatment the individual had received in

the intervening time which may have affected the outcome was also considered important. Completions of follow up of greater than 90% were thought to minimise verification bias. This was calculated from a flow chart on the data collection sheet detailing the number of eligible infants for each study, those that received the index test and outcome measure, and any reported exclusions or withdrawals from the study at each stage. The number or infants receiving the outcome measure/ number receiving the index test x 100 was used to calculate the completeness of follow up percentage.

The overall quality of a study was not used to define inclusion/ exclusion into a review, but, where possible, either meta-regression or subgroup analysis according to study quality was performed, and the overall quality of the included studies and adherence to each of the criteria described above are presented in table form in each review.

4.5 Data extraction

Information was extracted from the selected articles, in duplicate (by the author and a second reviewer as described in section 4.3) using a pre-designed data collection sheet. This was piloted for each review to ensure all relevant items were included. The data collection sheets for each review were similar and are given in Appendices 4 to 6. Data were extracted on study characteristics including population and setting, index test (including threshold values used), outcome measure (including blinding, definition and threshold), study quality and results, and were entered onto an Excel spreadsheet. Data were used to construct 2x2 tables of the association between the

index test of interest using the threshold reported in the paper and the postnatal outcome for each individual, as described in Table 4.1.

Table 4.1 2 x 2 table

	Outcome measure positive	**Outcome measure negative**	**Total**
Index Test positive	True positive (TP)	False positive (FP)	TP + FP
Index Test negative	False negative (FN)	True negative (TN)	FN + TN
Total	TP + FN	FP + TN	TP +FP+ FN+ TN

If results for multiple thresholds were reported, a separate 2x2 table for each threshold was constructed. In studies where data were felt to be relevant but 2x2 tables could not be completed, or the outcome or population reported in the paper did not meet the specific inclusion criteria, the authors were contacted. The study was not included unless the data could be provided and the population, and outcome, considered satisfactory. Difficulties in data extraction were resolved by seeking input from a third reviewer (Professor KS Khan).

4.6 Analysis and interpretation of data

In this thesis, the term ‘prognostic’ refers to strength of association between a test at birth and the odds of an adverse outcome, as measured by an odds ratio. The term ‘predictive’ refers to the ability of a test to discriminate between those babies who will and those who will not experience an adverse outcome, as measured by sensitivity,

specificity, and positive and negative likelihood ratios. A test may have strong prognostic ability, but not necessarily good predictive ability, and so it is important to consider both.[93]

4.6.1 Data synthesis for prognostic association

The 2 x 2 tables were used to compute odds ratios (OR) and 95 % confidence intervals (CI) for each index test-outcome pair. OR were calculated using the formula *(TP x TN)/ (FP x FN)*. The CI were calculated using the formula logOR +/- 1.96SE(logOR). ORs were selected as the summary statistic as they represent the effect of the exposure on the odds of having the condition in an unbiased fashion and enable the results of case- control and cohort studies to both be included.[94] It is frequently used to demonstrate an epidemiologic association,[94] and here it provides a measure of a test's prognostic ability. Where possible the results for each index test and outcome measure were pooled using meta-analysis.

4.6.2 Data synthesis for predictive ability

Where there was a strong and statistically significant prognostic association between a test and an outcome measure (defined by an OR >5 ,with 95% confidence interval that did not cross 1) sensitivity, specificity and likelihood ratios (LR) were calculated, again using data from the 2 x 2 tables as follows: sensitivity *TP / (TP+ FN)*; specificity *TN/ (FP + TN);* positive likelihood ratio *sensitivity/ (1-specificity);* negative likelihood ratio *(1-sensitivity)/ specificity*. This allowed the predictive ability of the test to be determined;[95] that is, whether the test can accurately discriminate between those who do and those who do not have a poor outcome (as measured by sensitivity and

specificity), and how much a positive or negative test result modifies the odds of a poor outcome (as measured by the positive and negative likelihood ratios).

4.6.3 Assessment of heterogeneity

OR data was plotted in forest plots and the between-study heterogeneity in the prognostic association for each test was assessed visually and by estimating I^2 (the amount of variability in prognostic effects due to between-study heterogeneity)[96] and tau-squared (an estimate of between study variance).[97]

Where significant heterogeneity (defined as I^2 >50%) was present, the reason for heterogeneity was explored using meta-regression (where the number of studies included in a particular meta-analysis was ≥ 10), planned *a priori* in keeping with published recommendations.[96] Where meta-regression was significant, or if meta-regression was not possible, subgroup analyses were performed to explore the effect on results and heterogeneity. Pre-defined categories included:

1. Study quality (high quality versus low or medium quality)
2. Population characteristics (including gestational age, year of birth of study population, risk factors for adverse outcomes)
3. Study setting (including country of origin where standards of care thought to be similar: USA/ Europe/ Australia and New Zealand versus others)

Factors thought to be relevant to particular index tests and their relationship with the outcome measures in question are reported in the relevant chapter.

4.6.4 Meta-analysis

Meta-analysis was performed where two or more studies used the same index test and outcome measure. In each study when a table contained cells with a value of 0, 0.5 was added to all cells to allow the calculation of log ORs and their variances for meta-analysis.[98] The primary outcomes were considered to be neonatal mortality, and composite measures of neonatal, childhood and adult morbidity. A composite outcome measure for morbidity was employed to maximise the number of events that could be included in the analysis and avoid the need to select a single morbidity as a primary outcome measure. However, a hazard of composite outcome measures is the assumption that the significance of the result applies to all components.[99] To address this issue, the component outcomes were analysed as subgroups where possible. When the composite outcome measure was used, care was taken to ensure that each individual was only counted once in each analysis, particularly where studies reported multiple outcomes for a single population. Where multiple outcomes were reported, attempts were made to select the outcome most consistent with other studies within the meta-analysis.

Due to the expected presence of clinical and statistical heterogeneity between studies, a random effects model was used throughout, which dichotomises the log OR estimates for each test and weights each study by the inverse of the study's variance plus between-study variance. When this method is used to calculate odds ratios, it provides a summary estimate of the average prognostic effect of a test.[100] As a test's prognostic ability may vary from this average from setting to setting, after each random-effects meta-analysis ,if I^2 was greater than 0%, a prediction interval (EPI) was calculated to reveal the potential prognostic association if the test is

applied in a single setting similar to one of the studies from the analysis.[101] This was calculated where three or more studies were included in the meta-analysis.

Pooled sensitivity, specificity and likelihood ratios were also calculated using a bivariate random-effects meta-analysis model. Bivariate meta-analysis accounts for the possible negative correlation between the sensitivity and specificity of a test, thought to be due to the fact that varying test threshold affects these parameters, and that different studies may have explicit or implicit differences in the threshold used to define a positive test. Explicit differences arise from studies using different thresholds to define a positive test, whereas implicit threshold variations may result from differences in observers or equipment.[102] Bivariate meta-analysis assumes that the sensitivities from different studies (after logit transformation) within a meta-analysis are normally distributed around a mean value, with variability around this, and the same considerations are applied to the specificities of the study. The combination of the two normally distributed outcomes, acknowledging the potential correlation between them, results in the bivariate normal distribution.[102]

All bivariate meta-analyses were performed in Stata version 10.0 (StataCorp, Texas, USA) using the metan and metandi commands.[103;104] Plots were generated using StatsDirect (StatsDirect, Cheshire, UK). Meta-Disc was used for calculations and non-bivariate meta-analyses.[105]

4.6.5 Publication bias

Publication bias arises when the studies included in a review differ systematically from those that are missed. Funnel plots assess sample size effects by plotting the log odds ratio (ln OR) against the sample size of precision (estimated by the

reciprocal of the standard error (SE). Where no sample size effect exists, the points will form a symmetrical funnel plot.[106] Sources of asymmetry include publication bias and location bias (e.g. language bias), poor methodological quality leading to spuriously inflated effects in smaller studies,[107] and true heterogeneity.[108] Certain tests of funnel plot asymmetry have been found to be more prone to type I error rates within meta-analyses of diagnostic accuracy, due to correlation between lnOR and its SE, and therefore these were avoided.[106] To explore for the presence of funnel plot asymmetry (small study effects) and thus potential publication bias, the Peters test was performed in each meta-analysis containing at least 10 studies.[108] This uses a weighted linear regression which has been shown to be more accurate.[109] The analysis was performed in Stata 10 using the metabias command.[110]

4.7 Description of data

For each neonatal test, data on individual studies are presented as follows:

1. Table of the characteristics of included studies including population, test and outcome measure.
2. Table of the methodological quality of the included studies according to the pre-specified criteria.
3. Forest plots of odds ratios of individual studies and summary OR from meta-analysis according to the index test, threshold and outcome of interest.
4. Sensitivity, specificity and likelihood ratios are presented in tabular form along with their 95% confidence intervals.
5. Forest plots or tables with subgroup analyses (where possible).

6. Results of the Peters test for publication bias (asymmetry) according to p value.

CHAPTER 5: SYSTEMATIC REVIEW AND META – ANALYSIS OF THE PROGNOSTIC AND PREDICTIVE ABILITY OF UMBILICAL CORD PH OR BASE EXCESS AT BIRTH FOR SHORT AND LONG TERM OUTCOMES

5.1 ABSTRACT

5.1.1 Background

The purpose of this systematic review was to evaluate the association between umbilical cord pH or base excess at birth and outcomes.

5.1.2 Methods

A systematic review of the literature with random effects meta-analysis and meta-regression, using standard techniques was performed. Odds ratios with their 95% confidence intervals were computed, and summary sensitivity, specificity and likelihood ratios, to assess predictive ability. Electronic searches were performed from database inception until August 2008 without language restrictions. The

reference lists of selected articles were screened and authors contacted. Studies were selected by two reviewers if umbilical cord pH at birth, and/or base excess, by any threshold, were related to neonatal or long term outcomes.

5.1.3 Results

51 articles including 481753 individuals met the selection criteria. Studies varied in design, quality, outcome definition and results. Meta-analysis performed within pre-defined groups showed that low arterial cord pH had significant associations with neonatal mortality (OR 16.4, 95% CI 8.9-30.4, I^2 0%), hypoxic ischemic encephalopathy (OR 15.2, 95% CI 7.0-33.0, I^2 0%), intraventricular hemorrhage or periventricular leukomalacia (OR 2.7, 95% CI 2.0-3.7, I^2 0%) and cerebral palsy (OR 2.1, 95% CI 1.3-3.4, I^2 0%). Umbilical arterial pH showed a high specificity but poor sensitivity for neonatal mortality and morbidity.

5.1.4 Conclusion

Low arterial cord pH showed strong, consistent and temporal associations with clinically important neonatal outcomes that are biologically plausible. These data can be used to inform clinical management and justify the use of arterial cord pH as an important outcome measure alongside neonatal morbidity and mortality in obstetric trials.

5.1.5 Publications arising from this work

Malin GL, Morris RK, Khan KS. Umbilical cord pH in the prediction of neonatal and long term morbidity: a systematic review of the literature and meta-analysis*. Archives of Disease in Childhood: Fetal and Neonatal Edition 2009; 94 (supplement 1): Fa 26*

Malin GL, Morris RK, Khan KS. What is the relationship between umbilical cord pH and perinatal and long term outcome? A systematic review and meta-analysis to evaluate strength of association. *BMJ 2010; 340: c1471.*

5.2 Introduction

Perinatal asphyxia is a major cause of neonatal and childhood morbidity and mortality.[111] It is predicated by fetal acidosis, determined by umbilical cord pH at birth.[112] Cerebral palsy is thought to occur more frequently at an arterial cord pH of <7.00 and a base deficit ≥ 12mmol/l.[71] However, these criteria have been derived through consensus, not through evaluation of collated evidence summaries in the field,[71] leading to clinical uncertainty.[72] As described in section 2.3.4, this is because existing observational studies of the association between cord pH and outcomes have drawn inconsistent inferences, and the only other systematic review in this area did not provide robust results.[77] Substantial uncertainty therefore remains about the value clinicians may attach to acidosis in the management of neonates, and the long term implications of a low arterial cord pH. A documented low umbilical cord pH is a factor which may be used to support medico-legal claims of harm during intra-partum events resulting in long term disability.[113] It is therefore imperative that the validity of this association is supported with high quality evidence.

The aim of this systematic review of the literature was to quantitatively establish the prognostic association and predictive ability of acidosis at birth with neonatal mortality, morbidity and long term outcomes, and assess if causal criteria were met.[114]

5.3 Methods

The methods employed are those outlined in Chapter 4, with those specific to this review detailed below.

5.3.1 Data sources and searches

Electronic searches were performed with the aim of capturing data regarding neonates with umbilical cord pH performed at birth, and adverse outcomes throughout the life course. Searches were performed by the author from database inception until August 2008. The search strategy employed in Medline is given in Appendix 7. This was adapted for use in other electronic databases.

5.3.2 Study selection

Studies were selected if they contained data on neonates who had an umbilical cord blood tested for arterial, venous pH or base excess at birth, and the association with neonatal or long term outcomes

5.3.3 **Data extraction and quality assessment**

The data extraction tool employed in this review is given in Appendix 4. Any threshold of arterial or venous pH, or base excess, taken from the umbilical cord after delivery was included. All live born neonates were included. All studies were assessed fully using the STARD and QUADAS checklists(Appendix 2 and 3).[90;91] The elements felt to be most relevant to systematic reviews assessing prognostic association and predictive ability, as described in Section 4.4, were used to assess the overall quality of the included studies. A study meeting four or more of the criteria was considered to be of high quality, three moderate, and two or less of low quality.

5.3.4 Data synthesis

The 2 x 2 tables were used to compute odds ratios (OR) and 95 % confidence intervals (CI) for each index test-outcome pair, and pooled the results for each index test (considering each definition and threshold of growth as a separate test) using meta-analysis.

Analyses were performed for groups defined *a priori* according to index test (arterial cord pH, venous cord pH and base excess). The main outcome measures considered included neonatal mortality, a composite measure of neonatal morbidity, and cerebral palsy. The component neonatal outcomes of hypoxic ischemic encephalopathy, seizures, intraventricular hemorrhage or periventricular leukomalacia were analysed separately. Within the composite analysis of neonatal morbidity, HIE was the most commonly reported across the included studies, and therefore this outcome was selected if studies reported more than one measure of morbidity.

Within the largest groups, (arterial cord pH paired with neonatal mortality and neonatal morbidity), meta-regression was performed to explore the reasons for heterogeneity. Study design (cohort versus case-control), study quality (high versus medium/ low quality) and population risk (high versus low/ unselected or unreported) were considered potential sources of heterogeneity. Population at high risk for complications was defined based on reported characteristics including CTG abnormalities, meconium liquor, low Apgar score at birth, gestation < 37 weeks, and birth weight < 2500g. Comparison of a low risk population to others was not possible as only one included study met the criteria for this definition. Exploration of other

potential sources of heterogeneity was not possible given the reporting quality of the primary studies. Multivariable and univariate meta-regression were performed, in accordance with published guidance, which recommends allowing at least 10 studies per covariate explored.[38] Where a source of heterogeneity was identified, subgroup analyses were performed. As complications such as cerebral palsy and intra-ventricular haemorrhage have been shown to be increased in pre-term and low birth weight infants, studies with these populations were analysed separately, and compared the results to that of a term population (≥ 37 weeks gestation).

In order to assess causal association for each outcome Hill's criteria were considered.[115;116] To explore the prognostic association and predictive accuracy of umbilical arterial cord pH at different thresholds, the summary odds ratios and sensitivity and specificity were calculated for subgroups of studies using different thresholds to predict neonatal mortality and morbidity.

5.4 Results

5.4.1 Literature identification and study characteristics

As shown in Figure 5.1, after an initial search of 5690 citations, 51 primary articles were included, two following provision of further information from authors. 481753 individuals were included in the overall review. Some of the studies reported multiple umbilical cord parameters or outcomes; these are counted in each relevant category but only once in the overall total. 43 of the studies were eligible for inclusion in meta-analyses according to pre-defined outcome measures, including 479383 individuals in total and 70 2x2 tables**.** The characteristics of included studies are given in Appendix 8, and the references in Appendix 9.

5.4.2 Data extraction and quality assessment

The majority of papers reported arterial cord pH (n=46), with thresholds ranging from 7.00-7.24. Arterial base excess was reported in five studies, with thresholds of 12 to 16mmol/l. Four studies reported venous cord pH; three using a threshold of 7.20 and one 7.10. Only one study utilised venous base excess. The umbilical cord pH level was always obtained prior to the occurrence of the adverse outcome. A wide variety of outcome measures were reported, ranging from neonatal mortality and morbidity to long term outcomes including cerebral palsy, unspecified neurological abnormality, intelligence quotients and developmental assessments including the Bayley and Griffiths scores. Where the same outcome was reported the thresholds and ages at assessment varied between studies. 36 studies reported only on outcomes within the neonatal period. 15 studies performed long term follow up, one did not report the age of ascertainment of the outcome. The range reported in other studies was 1 to 8 years; the median across studies was 5 years.

Overall study quality was variable (Table 5.1). The majority of studies were retrospective but with a cohort design, allowing inferences concerning temporality of association. Over 80% of studies met the following quality items: appropriate outcome measure, > 90% verification with outcome measure, and cohort design. Studies scored poorly on the following items: adequate description of index test and outcome measure, and consecutive recruitment. All included studies were unclear regarding the use of medical intervention between the index test and outcome measure. Overall 45% of studies were high quality, 39% medium and 16% low. Study design and quality did not appear to influence results on meta-regression (Table 5.2).

5.4.3 Relationship between cord blood assessment at birth and outcomes

Association between cord blood assessment at birth and neonatal mortality

There were 15 studies (13 cohort design and 2 case control design) with a total of 469395 infants, which reported the association between arterial cord pH and mortality. All studies had an OR point estimate > 1.0. There was significant heterogeneity (I^2 61.0%) overall. Meta-regression (Table 5.2) identified population risk as a significant explanatory factor. Within sub-groups (Figure 5.2) the association was consistent across studies. The association of low umbilical artery pH with neonatal mortality was stronger in the unselected population (OR 16.9, 95% CI 9.7 to 29.5, I^2 0%) than in the high risk population (OR 4.3, 95% CI 2.2 to 8.5, I^2 29.5%). The analysis for the unselected population was dominated by one very large study, which was based on data from a regional perinatal register where umbilical cord pH was performed at the majority of births (Heller et al 2003). A sensitivity analysis was performed excluding this study. This did not significantly alter the point estimate (OR 17.0, 95% CI 4.4 to 65.5), however the estimated prediction interval became very wide (0.0 to 106299.0)

Examination of the high risk population further showed that studies reporting on a population of infants born at less than 32 weeks gestation, or with a birth weight <2,000g, when analysed separately (7 studies), had a significant association between cord pH and mortality (OR 3.5, 95% CI 2.3 to 5.4, I^2 0%). When limited to a term (>37 weeks gestation) population (4 studies), the association was also strong (OR 9.3, 95% 1.4 to 63.2, I^2 84%). Grouping the studies according to quality did not affect the significance or direction of association of the pooled result (Figure 5.2).

Exploration for a threshold effect (Table 5.3) showed that the results for a cut off of pH 7.00 the association did not reach significance overall (OR 6.1, 95% CI 0.9 to 41.6) and the prediction interval was very broad (0.0 to 20406.6). The results for a threshold of 7.10 gave a similar point estimate but achieved significance (OR 7.1, 95% CI 3.3 to 15.3); however the predictive interval remained broad (0.8 to 64.3) and crossed the line of no effect. For a threshold of 7.20, the OR was lower (4.3, 95% CI 2.2 to 8.7) with a broad predictive interval (0.5 to 40.6). Only one study (Heller et al 2003) examined all 3 thresholds, which demonstrated the strongest association at threshold 7.00 (OR 16.9, 95% CI 9.2 to 31.1) and the weakest at 7.20 (OR 3.1, 95% CI 2.3 to 4.1).

Association between cord blood assessment at birth and neonatal morbidity

There were 31 studies comparing umbilical artery pH with a variety of neonatal outcomes (Figure 5.3). One of these was excluded from meta-analysis of composite neonatal morbidity because it contained a duplicate population to another included study, therefore the dataset to assess the association of arterial pH with a composite measure of morbidity included 30 studies with 10904 individuals. Only two studies had an OR point estimate < 1.0, the rest showed an association. However, significant heterogeneity was present (I^2 58.2%). Meta-regression showed population to be an explanatory variable (Table 5.2). Subgroup meta-analysis for a high risk population (Figure 5.3) showed a weaker association (OR 3.4, 95% CI 2.3 to 4.9) than in an unselected/ undefined population (OR 10.6, 95% CI 4.7 to 24.1, I^2 66.4%). When analysed in subgroups according to quality, the direction of effect remained consistent and significant between the group of 12 high quality studies (OR 6.6, 95% CI 3.7 to 11.8, I^2 51.2%) and 18 low/medium quality studies (OR 4.6, 95% CI 2.5 to

8.4, I^2 61.5%). Exploration by subgroup analysis for a threshold effect (Table 5.3) showed the most substantial association exists at a threshold of 7.00 (OR 12.5, 95% CI 6.1 to 25.6). The EPI was broad but did not cross the line of no effect (1.7 to 89.9) The results for a threshold of 7.10 (OR 2.4, 95% CI 1.3 to 4.3) and 7.20 (OR 2.2, 95% CI 1.3 to 3.7) were similar.

When components of the composite outcome were analysed (Figure 5.3), the association between umbilical artery pH and HIE had an OR of 13.8 (95% CI 6.6 to 28.9, I^2 0%). That between arterial pH and seizures had an OR of 8.1 (95% CI 3.0 to 21.9, I^2 66.3%). A low arterial pH was associated with IVH or PVL (Figure 5.3) with an OR of 2.9 (95% CI 2.1 to 4.1, I^2 0%). Only 2 of 9 studies reporting this outcome were not limited to a pre-term (<32 weeks) or low birth (<2000g) population. Excluding these 2 studies did not affect the strength or significance of the association.

The association between venous pH and neonatal morbidity was reported in 5 studies. Meta-analysis demonstrated a significant association (OR 2.9, 95% CI 1.9 to 4.4, I^2 44.5%). Four studies examined the association between arterial base excess and neonatal morbidity, which was similar to venous pH (OR 2.6, 95% CI 1.4 to 4.9, I^2 0%). Due to the small number of studies reporting this variable, it was not possible to combine the cord pH and base excess to compare the difference between respiratory and metabolic acidosis.

Association between umbilical cord pH at birth and cerebral palsy

Seven studies examined the association between umbilical artery pH and cerebral palsy, including 1117 individuals. Of those, two had an OR point estimate <1.0, but

the rest showed an association of low pH with cerebral palsy (Figure 5.4). The criteria used to diagnose cerebral palsy were given in only two studies. The overall association was OR 2.3 (95%CI 1.3 to 4.2, I^2 0%). It was not possible to explore the threshold effect by subgroup analysis due to the small number of studies reporting this outcome.

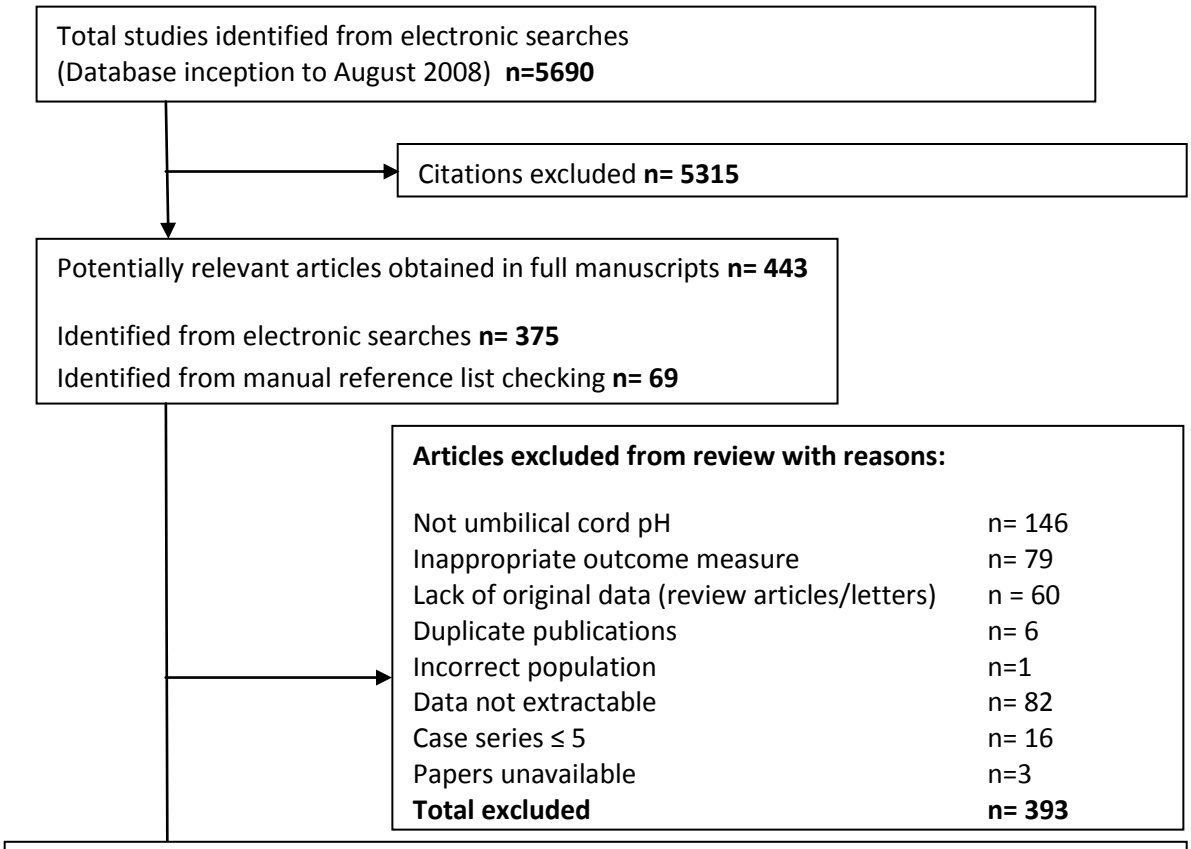
Predictive ability of umbilical cord pH at birth for neonatal morbidity and mortality

As reported in Table 5.3, at the lowest threshold of umbilical arterial pH reported (pH <7.00), specificity was high for neonatal mortality and morbidity (1.0, 95% CI 0.99 to 1.0 and 0.92, 95% CI 0.91 to 0.93 respectively) but sensitivity low, indicating that neonates with a positive test (i.e. a low cord pH) are at high risk of adverse outcome, but that a negative test does not change the odds of a poor outcome significantly. Increasing the threshold to 7.20 reduced the specificity and slightly increased the sensitivity, but not to a level where it would be considered a good discriminator for 'test negative' infants.

5.4.4 Publication Bias

Within the group of studies considering the association between arterial pH and neonatal mortality (n=15), Peters test was not significant (p=0.318). For neonatal morbidity (n=31), the result was also not significant (p=0.847) indicating no small study effects.

Figure 5.1: Study selection process for systematic review of umbilical cord pH or base excess and neonatal and long term morbidity and mortality



Selected for inclusion in systematic review n= 51 (481753 patients)

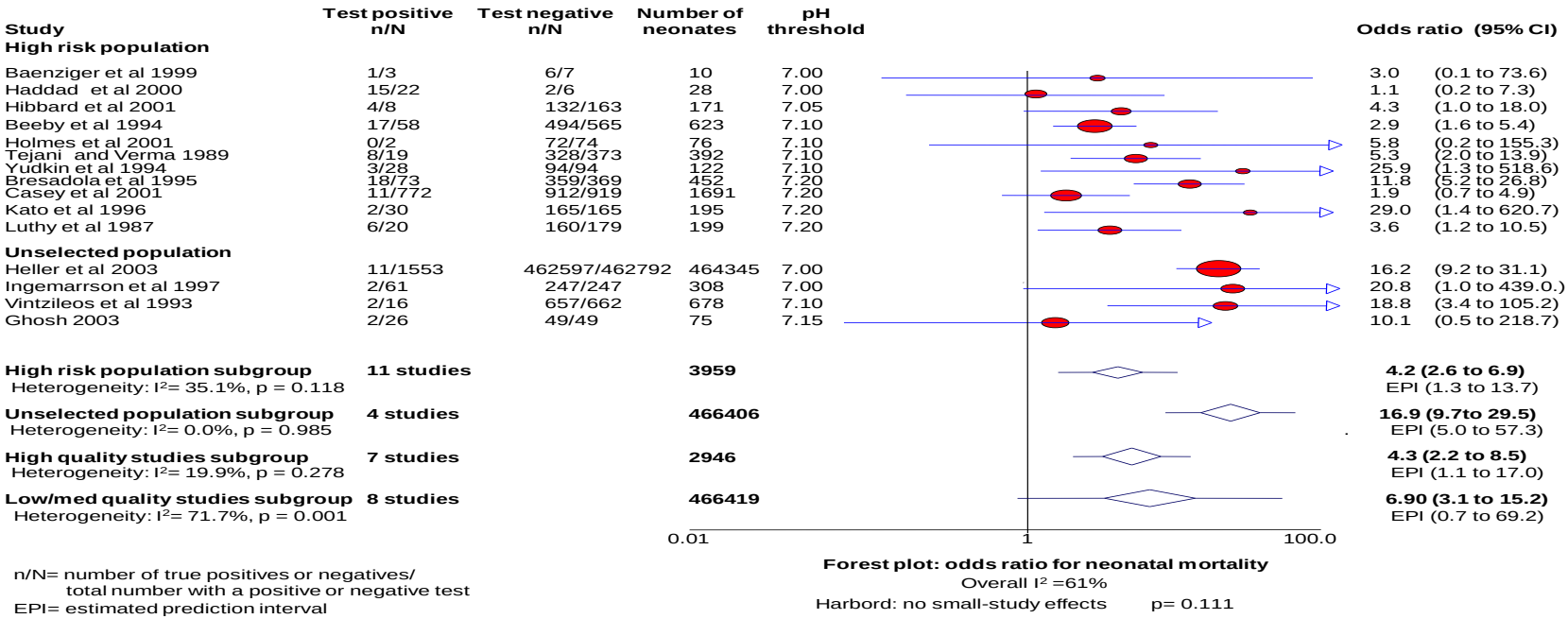
Test	No. of studies	No. of patients
Arterial cord pH	46	479022
Venous cord pH	6	2531
Arterial base excess	4	549
Venous base excess	1	25

Total primary articles included evaluating arterial cord pH n=46 (479022 patients)

Outcome measure	No. of studies	No. of patients
Neonatal mortality*	15	469365
Composite neonatal morbidity*	30	10904
– **Hypoxic ischaemic encephalopathy***	7	827
– **Seizures***	9	5191
– **Intraventricular haemorrhage/periventricular leukomalacia***	12	284
Cerebral palsy*	7	1117
Comprehension and reading ability	1	116
Intelligence	2	41
Speech/language	3	423
Global development	3	359
Movement disorder	1	44

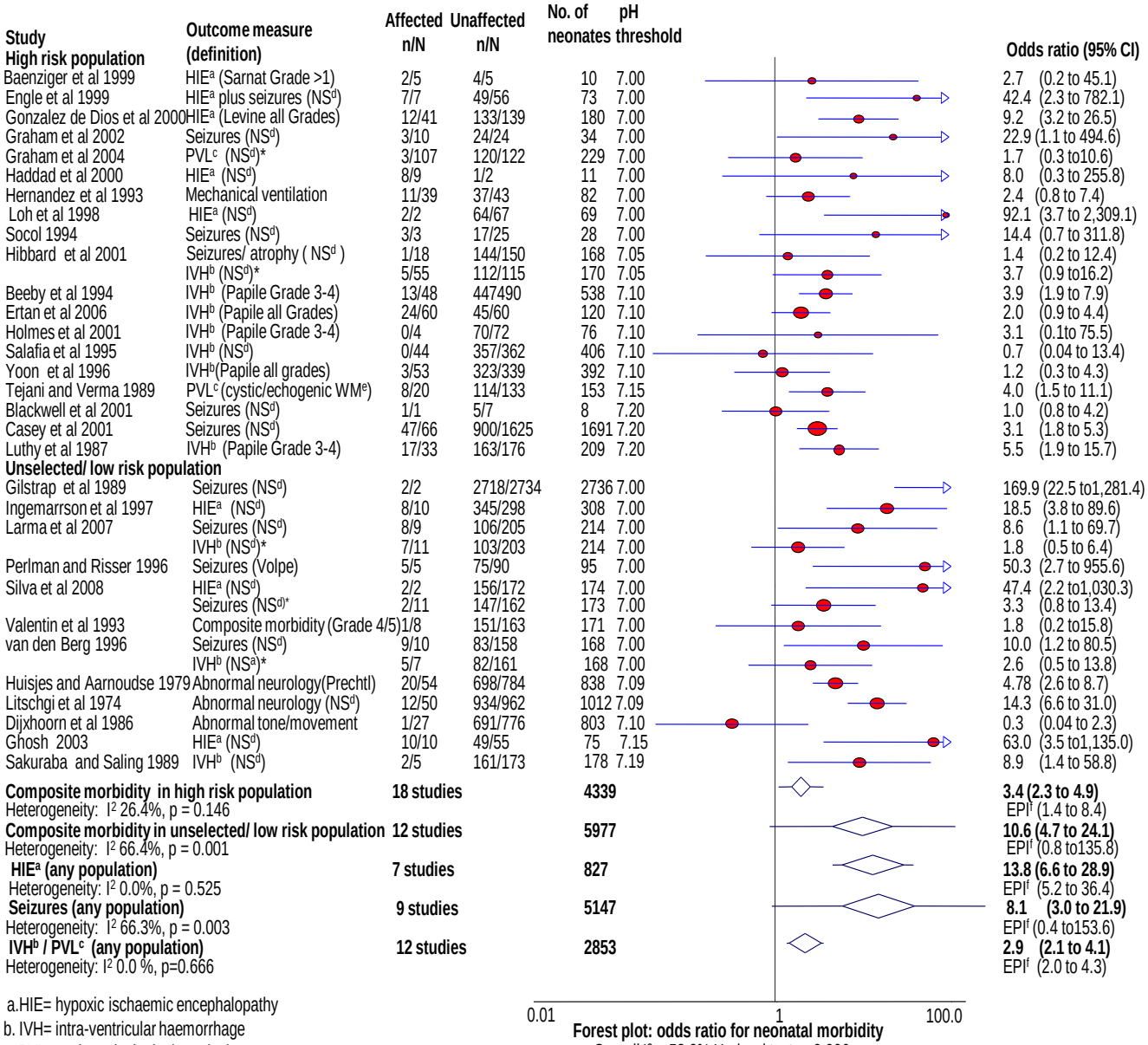
*pooled in meta-analyses

Figure 5.2 Forest plot of odds ratios for the association of umbilical arterial cord pH with neonatal mortality. Single studies are represented by a filled circle, and pooled studies (and 95% confidence intervals) by a diamond



n/N= number of true positives or negatives/
total number with a positive or negative test
EPI= estimated prediction interval

Figure 5.3 Forest plot of odds ratios for the association of umbilical arterial cord pH with neonatal morbidity. Single studies are represented by a filled circle, and pooled studies (and 95% confidence intervals) by a diamond



a.HIE= hypoxic ischaemic encephalopathy

b. IVH= intra-ventricular haemorrhage

c. PVL= periventricular leukomalacia

d. NS= not stated

e. WM= cerebral white matter

f. EPI= Esti

*Not included in 'any morbidity' meta- analysis

•n/N refers to the number of true positive or true negative/ total affected or unaffected

Figure 5.4 Forest plot of odds ratios for the association of umbilical arterial cord pH with cerebral palsy. Single studies are represented by a filled circle, and pooled studies (and 95% confidence intervals) by a diamond

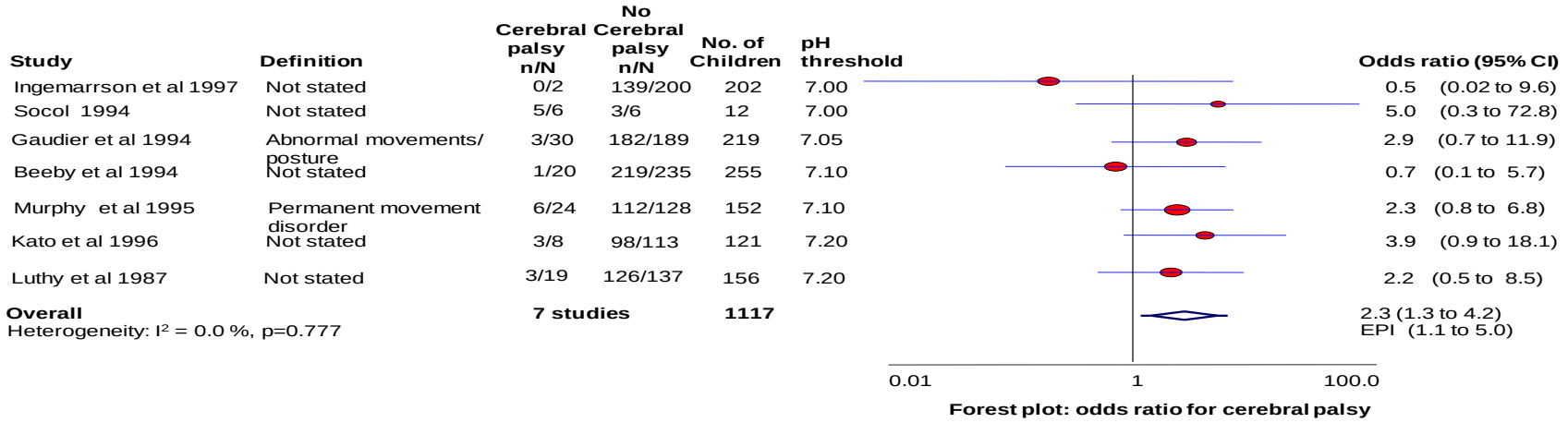


Table 5.1 Methodological quality of studies included in systematic review of umbilical cord pH and neonatal and long term outcomes

	Number (%) of studies n= 51		
Quality Item	**Yes**	**No**	**Unclear**
Cohort study design	42 (82)	9 (18)	0
Population adequately described	40 (78)	2 (4)	9 (18)
Consecutive recruitment	2 (4)	1 (2)	48 (94)
Prospective recruitment	15 (29)	29 (57)	7 (14)
Appropriate outcome measure	51 (100)	0	0
Outcome measure blinded	7 (14)	0	44 (86)
>90% of individuals had outcome measure	41 (80)	10(20)	0
Index test and outcome measure described	8 (16)	12 (24)	31 (61)
Intervention between index test and outcome	0	0	51 (100)
Quality Classification			
High	23 (45)	-	-
Medium	20 (39)	-	-
Low	8 (16)	-	-

Table 5.2 Exploration of heterogeneity in the estimation of association of low arterial cord pH with neonatal mortality and morbidity

	Univariable analysis		Multivariable analysis	
Factor	Odds ratio (95% CI)	p value	Odds ratio (95% CI)	p value
Outcome: Neonatal mortality				
Study design: cohort v. case control	0.4 (0.03-6.5)	0.52	0.8 (0.1-13.1)	0.87
Study quality: high v. medium or low quality	0.8 (0.3-2.4)	0.63	0.7 (0.3-2.1)	0.52
Population: high risk v. unselected	0.3 (0.1-1.0)	0.048	0.3 (0.1-1.1)	0.06
Outcome: Composite neonatal morbidity				
Study design: cohort v. case control	1.1 (0.2-5.3)	0.90	1.0 (0.3-3.2)	0.98
Study quality: high v. medium or low quality	1.4 (0.5-3.7)	0.51	1.6 (0.6-3.8)	0.32
Population: high risk v. unselected or low risk	0.4 (0.2-0.9)	0.03	0.4 (0.2-0.9)	0.02

Multivariable analysis including quality grade, study design, and population risk as explanatory variables. See sections 4.6.3 and 5.3.4 for details.

Univariable analysis using dummy variables to set the reference category as medium/low quality, case- control design and unselected/ low risk population.

Table 5.3 The effect of the use of varying thresholds of umbilical artery pH on the association and predictive ability of arterial cord pH for neonatal morbidity and mortality

Outcome measure	Number of Studies	Umbilical cord pH threshold	Sensitivity (95% CI)	Specificity (95% CI)	Odds Ratio (95% CI)
Neonatal mortality	4	7.00	0.13 (0.09-0.18)	1.0 (0.99-1.0)	6.1 (0.89-41.61)
	6	7.10	0.19 (0.15-0.23)	0.98 (0.98-0.98)	7.09 (3.3-15.26)
	5	7.20	0.39 (0.33-0.45)	0.85 (0.85-0.85)	4.31 (2.15-8.67)
Neonatal morbidity	15	7.00	0.51 (0.43-0.59)	0.92 (0.91-0.93)	12.47 (6.08-25.59)
	10	7.10	0.24 (0.20-0.29)	0.92 (0.91-0.92)	2.38 (1.22-4.65)
	6	7.20	0.55 (0.48-0.63)	0.62 (0.60-0.64)	2.41 (1.42-4.08)

5.5 Discussion

Low arterial cord pH had a strong, consistent and temporal association with neonatal mortality and morbidity (composite of hypoxic ischemic encephalopathy, seizures and intraventricular hemorrhage / periventricular leukomalacia) and long term outcome (cerebral palsy). In all of the associations between arterial pH and outcome explored, with the exception of composite morbidity in a low risk population and seizures, the estimated prediction interval suggested that a future study would have the same direction and significance of effect observed. The associations observed are biologically plausible.[76, 117-119]

Strengths and limitations of this review

The strength of this review and the validity of the inferences lie in the methodology used. It complies with existing guidelines for the reporting of systematic reviews of diagnostic and observational studies evaluating causal association.[78;116] The most up to date techniques for performing and interpreting meta-analysis were used.[84;85] An extensive literature search was performed in relevant databases with no language restrictions applied. The pooled studies had heterogeneity in terms of quality, population risk, threshold of umbilical cord pH used, and ascertainment of neonatal outcome. Recommended analyses were performed to address this issue, including bivariate and subgroup meta-analysis to take into account the threshold effect,[102] meta-regression analysis to explore for reasons for heterogeneity,[120] and component outcome analysis to examine the suitability of the composite morbidity outcome. This did not significantly affect the results. The observed associations were qualitatively in the same direction and statistical heterogeneity, where present, arose from variation

in strength of associations from study to study rather than opposition in direction of association. The inferences concerning a causal association between low arterial cord pH at birth and neonatal death and a variety of neonatal morbidities therefore merit consideration.

There are several limitations to this review. The quality of the primary studies was variable. Meta-regression and subgroup analysis were employed to explore for the effect of this issue, which demonstrated that limiting to high quality studies did not affect the results. The poor reporting of population characteristics limited the subgroup analysis according to risk factors: a large number of the papers with an 'unselected' population did not fully report characteristics such as birth weight or gestational age, which means that it is difficult to extrapolate the findings to the general obstetric population. Only one paper included in the meta-analysis specified that it was limited to a low-risk term population. With regard to the index test examined, only a small number of papers reported base excess, which meant that no comparison of metabolic and respiratory acidosis was possible. The results demonstrated that a high base excess is associated with neonatal morbidity, but in clinical practice the pH and base excess level would be considered together in any individual, and this could not be commented upon. Additionally, the analysis could only be based on the thresholds reported in the primary studies, which limited exploration of the effect of varying pH levels. Some of these issues may be dealt with using individual patient data meta-analysis.[121]

Venous cord pH and arterial base excess showed weaker associations with neonatal morbidity than that of arterial cord pH, with estimated predictive intervals that crossed the line of no effect. This analysis was however based on a small number of studies

and no other subgroup analysis could be performed, therefore a direct comparison with arterial cord pH was not possible. It seems likely from the results that arterial pH shows a stronger association with outcome.

Evidence to support the causality of association of a low umbilical artery pH with adverse outcomes

Hypoxia-ischemia initiates energy depletion, accumulation of extracellular glutamate and activation of receptors, leading to a deleterious cascade of events resulting in neuronal death.[11] However, different areas of the brain are susceptible to injury at different stages in development and the consequences of injury are unpredictable.[16] This would support the findings of a strong association with hypoxic ischemic encephalopathy but a weaker association with cerebral palsy. Only 10% of infants with evidence of hypoxic ischemic encephalopathy develop cerebral palsy and the reasons for this are not yet fully understood.[122] This issue is further highlighted by the findings of the term breech trial, where long term follow up did not reveal any difference in neurodevelopmental delay, despite an apparent increase in neonatal morbidity in the vaginal breech group.[123]

It is difficult to comment on the specificity with which umbilical cord pH is associated with the outcomes examined within the review. All of the outcomes examined may arise from a variety of causes, for example neonatal seizures may be caused by congenital brain anomaly, infection and metabolic disorders in addition to hypoxia. Although some of the included studies excluded cases of congenital anomaly, others did not specify that they had excluded individuals with an adverse outcome related to another cause from their analysis. The strength of association of a low arterial cord

pH with neonatal mortality and morbidity was stronger in the unselected than high risk group. This may be explained by the fact that individuals in the high risk group are more likely to suffer an adverse outcome from a cause other than asphyxia (e.g. prematurity, low birth weight), therefore although there is a proportionally higher number of deaths in the high risk group, there is a higher specificity for low arterial cord pH and death in the unselected population. Due to the fact that the primary studies did not report the cause of death within the study population, it is not possible to compare the odds ratios for death related to asphyxia, or subdivide the high risk population further.

The strength of association and predictive ability was calculated for different pH thresholds, between arterial cord pH and neonatal mortality and morbidity. Although the meta-analysis did not show a clear dose- response relationship, and the estimated predictive intervals were broad, the association for both neonatal mortality and morbidity were weakest at the highest threshold. Only one study (Heller et al 2003) explored the association of arterial pH and neonatal mortality at all three thresholds examined. Within this study, a clear dose- response relationship with the strongest association at a threshold of 7.00 and the weakest association at 7.20 was apparent. The within study comparison may be more valid as it avoids confounding by other study level factors which may affect between-study comparisons.

Implications for clinical practice

Umbilical cord pH is currently performed in infants believed to be at high risk for neonatal asphyxia. However, the results suggest that the strength of association with cord pH and outcome is not limited to this population group. Therefore future

research should assess the use of cord pH across neonatal populations, particularly exploring the cost effectiveness of performing the test on all neonates.

Based on the review findings, increased initial surveillance of neonates born with a low arterial pH, regardless of their clinical condition, is warranted as the odds of complications have been shown to be higher in this group. The avoidance of a low pH at birth should continue to be a target for day to day obstetric practice. The results justify the use of arterial cord pH as an important outcome measure alongside neonatal morbidity and mortality in obstetric clinical trials. It is difficult to draw strong conclusions regarding the necessity of long term follow up for babies with low arterial cord pH, as the observed association with cerebral palsy was in a limited number of primary studies and the association, though statistically significant, was only moderately strong, and based on studies with varying quality.

Recommendations

Further research is required to clarify the optimum threshold and population in which umbilical cord pH should be performed to predict adverse outcome. A recent review highlighted the current deficits in prognosis research, and suggested that individual patient data (IPD) meta-analysis, where high quality primary studies exist, enables the prognostic value of a test to be assessed at an individual level.[124] However, there were a lack of high quality primary studies within this field which would impair IPD analysis. A large prospective cohort study with long term follow up, accounting for potential confounding factors, is therefore required. Such a study must include evaluation of outcomes relevant to the individual and society, such as developmental impairment and use of health care and educational resources, before the use of cord

pH as a prognostic test and predictive test, and its cost effectiveness, may be clarified. This will enable further exploration of the threshold effect and the use of the combination of pH and base excess to predict outcome. Future smaller studies should adhere to the STARD reporting criteria to facilitate meta-analysis.

5.6 Conclusion

A low umbilical arterial pH at birth is significantly associated with neonatal mortality, morbidity and cerebral palsy. Test positive (low pH) babies are at a substantially increased risk of neonatal mortality and morbidity, especially at a pH of <7.00. However, babies who test negative do not have a decreased risk of a neonatal mortality or morbidity. Further research is required to define the population in which umbilical cord pH testing should be performed and the threshold which should be used in clinical practice.

CHAPTER 6: SYSTEMATIC REVIEW AND META – ANALYSIS OF THE PROGNOSTIC AND PREDICTIVE ABILITY OF CURRENT BIRTH WEIGHT STANDARDS FOR SHORT AND LONG TERM OUTCOMES

6.1 Abstract

6.1.1 Background

The purpose of this systematic review was to evaluate whether different birth weight standards are associated with subsequent outcome in infants born at term (≥ 37 weeks gestation).

6.1.2 Methods

Systematic review of the literature, with random effects meta-analysis, to compute summary odds ratios (OR) to assess prognostic association, and summary sensitivity, specificity and likelihood ratios, to assess predictive ability.

Electronic searches were performed from database inception until January 2011 without language restrictions. The reference lists of selected articles were screened and authors contacted. Studies were selected by 2 reviewers if growth restriction at

birth, by any method and threshold, was related to neonatal or long term outcomes. The population was limited to infants born at term (≥ 37 weeks gestation) to avoid the confounding effects of prematurity.

6.1.3 Results

92 articles including 23051341 individuals were selected. Meta-analysis performed according to definition of growth restriction showed that absolute birth weight was strongly associated with neonatal mortality, with a birth weight <1.5kg giving the largest prognostic effect (summary OR 48.6, 95% CI 28.62 to 82.53) and increasing thresholds (2.0kg, 2.5kg, and 2.9kg) reducing this association but remaining high (2.5kg: summary OR 8.46, 95% CI 6.25 to 11.46). When using centile charts to define low birth weight, regardless of the threshold chosen, the summary ORs were also highly significant for neonatal mortality, but were closer to 1 than when using an absolute birth weight of 1.5kg or 2.0kg (<10^{th} centile summary OR 4.11, 95% CI 3.70 to 4.56). The association between birth weight standards and childhood and adult morbidity varied; there was no significant relationship between birth weight <2.5kg (summary OR 1.33, 95% CI 1.00 to 1.78) or population chart <10^{th} centile (summary OR 1.17, 95% CI 0.93 to 1.48) with a composite measure of morbidity in either age group. For all tests, summary predictive ability was generally a high specificity and a high positive likelihood ratio for neonatal death, but a low sensitivity, and a negative likelihood ratio close to 1, indicating that being test positive substantially increases the risk of neonatal mortality, but being test negative does not reduce the risk.

6.1.4 Conclusion

Absolute birth weight is a prognostic factor for neonatal mortality as it is strongly associated with this outcome, with the strongest associations at the lowest thresholds. Centile charts or other definitions of low birth weight at a variety of thresholds are also prognostic, but the indirect evidence suggests they are not as strongly associated as absolute birth weight. The association between low birth weight and childhood and adult morbidity was inconsistent. In terms of predictive ability at the individual-level, test positive babies are at substantially increased risk of neonatal mortality, but being test negative generally does not reduce the risk of death in the neonatal period.

6.1.5 Publications arising from this work

Malin GL, Morris RK, Riley RD, Teune M, Khan KS. Should we forget about centile charts? Comparing definitions of fetal growth restriction to predict adverse outcome. *Archives of Disease in Childhood: Fetal and Neonatal Edition 2012; Suppl 1). A111.*

6.2 Introduction

The 'Fetal origins' hypothesis suggests that malnourishment in utero changes fetal programming, whereby biological pathways are altered, resulting in increased susceptibility to disease. In 1986 Barker et al demonstrated an inverse relationship between birth weight and adult cardiovascular disease.[39] Since then, numerous studies have evaluated the association between low birth weight and morbidity and mortality from the neonatal period through to late adulthood.[36;37;42-48;125] However, the results have not always been consistent.[38;49-51]

A number of methods have been used to identify babies at risk of adverse outcome, including population based centile charts, the most commonly used threshold being the tenth centile;[126] customised charts where the mother's BMI and ethnicity are used to calculate individualised growth centiles;[127] and ponderal index which takes into account the neonatal weight and length.[128] The published associations between each standard for defining being small for gestational age and adverse outcome vary, and there is no current consensus regarding the best method.[129]

The aim of this systematic review was to re-examine the association between low birth weight and adverse outcomes, avoiding the confounding influence of prematurity, and to determine which definition of small for gestational age has the strongest prognostic and predictive association with subsequent morbidity and mortality from the neonatal period through to adulthood.

6.3 Methods

The methods employed are those outlined in Chapter 4, with those specific to this review detailed below.

6.3.1 Data sources and searches

Electronic searches were performed with the aim of capturing neonates with any birth weight standard, used to define being small for gestational age, and adverse outcomes throughout the life course. Searches were performed by the author from database inception until January 2011. The search strategy employed in Medline is given in Appendix 10. This was adapted for use in other electronic databases.

6.3.2 Study selection

Studies were selected if they examined the association between a measure of low birth weight and an adverse outcome in neonates born alive at ≥ 37 weeks gestation.

6.3.3 Data extraction and quality assessment

The data extraction tool employed in this review is given in Appendix 5. Any birth weight standard considered to be a measure of small for gestational age at any threshold was included (e.g. absolute birth weight 1.5, 2.0 and 2.5kg; population birth weight chart <10th centile). For population, studies had to clearly state that the neonates included were born at term (≥ 37 weeks gestation). If this was unclear, or if a mixed population was reported, authors were contacted to ask for clarification or to provide data relating to term born infants only.

All studies were assessed fully using the STARD and QUADAS checklists (Appendix 2 and 3).[90;91] The elements felt to be most relevant to systematic reviews assessing prognostic association and predictive ability, as described in Section 4.4, were used to assess the overall quality of the included studies. A study meeting four or more of the criteria was considered to be of high quality, three of moderate quality, and two or less of low quality.

6.3.4 Data synthesis

The 2 x 2 tables were used to compute odds ratios (OR) and 95 % confidence intervals (CI) for each index test-outcome pair, and the results pooled for each index test (considering each definition and threshold of growth as a separate test) using meta-analysis.

Summary OR data is presented in forest plots. Meta-analyses were performed where 2 or more studies reported the same index test and outcome measure. The outcome measures were grouped according to neonatal (up to 28 days), infant (up to 1 year), child and adolescent (up to 18 years) and adult (over 18 years). The primary outcomes were considered to be neonatal neonatal mortality (up to 28 days), and a composite measure of morbidity in infancy (up to 12 months), childhood and adolescence (12 months to 18 years) and adulthood (>18 years).

When the composite outcome measure was used, attempts were made to select the most consistent threshold and outcome across the analysis, for example in the childhood morbidity analysis hypertension was the most commonly reported outcome therefore this was selected primarily, followed by other components of the metabolic syndrome.

Within the largest meta-analyses, subgroup analysis was performed where possible to examine the effect of potential confounding factors. Singleton or multiple birth status, ethnicity, exclusion of congenital anomalies and birth of the study population during or after 1990 (due to recent advances in antenatal and neonatal care), and study quality were considered to be important factors which may influence the strength of the association between low birth weight and adverse outcome.

For the purposes of the meta-analyses, data where birth weight had been dichotomised around a threshold specified in the primary studies was used. In order to compare the effect of birth weight as a continuous variable, all of the included studies where logistic regression analysis had been performed were examined and the findings qualitatively summarised.

6.4 Results

6.4.1 Literature identification and study characteristics

As shown in Figure 6.1, after an initial search of 36956 citations, 92 primary articles were included, 25 after contact with authors who provided data or information. 23051341 individuals were included in the review in total. Some of the studies reported multiple birth weight parameters; these are counted in each relevant category but only once in the overall total. Characteristics of the included studies are given in Appendix 11, and the references of studies included in this review are presented in Appendix 12.

A total of 145 further articles were felt to contain potentially relevant data, but either the authors could not be contacted, could not supply data to create 2 x 2 tables, or on clarification regarding the population the study was excluded. If a study included infants of less than 37 weeks gestation, it was only included if separate data regarding term infants was given or the authors provided this. A number of studies contained duplicate populations with each other: where there was duplication of the test and outcome measure the least complete study was excluded from the review. If the population was the same but the measure of birth weight or adverse outcome differed both studies were included, but care was taken not to count any individual twice within a single meta-analysis, or within the overall numbers.

The majority of studies used population growth chart <10th percentile (n=34) or birth weight <2.5kg (n=31) as the index test. A wide variety of outcome measures including mortality and morbidity (e.g. hypertension, diabetes mellitus, learning difficulties, cerebral palsy) were reported.

6.4.2 Study quality assessment

The results for the quality assessment are presented in Table 6.1. The majority of included studies were of cohort design (95%), most were retrospective studies (58%). Most studies were of high or moderate quality according to the pre-specified criteria. Studies often failed to adequately describe the test or outcome in a way that would make them reproducible, and very few studies described any interventions that were performed between the time of the birth weight measurement and the outcome test. Where possible a subgroup analysis using only high quality studies was performed and the results are presented in Table 6.2. No subgroup analysis for study quality was performed for population <10th centile and adult outcomes, as only one of the studies was considered to be of high quality.

6.4.3 Data analysis

Neonatal outcomes

A forest plot of the summary meta-analysis odds ratios and 95% confidence intervals for each measure of being small for gestational age and neonatal mortality is given in Figure 6.2. Birth weight <1.5kg showed the strongest association with neonatal mortality (OR 48.6, 95% CI 28.62 to 82.53), with no between-study heterogeneity in this effect. Raising the birth weight threshold to 2.0kg, 2.5kg or 2.9kg gradually reduced the association and increased the heterogeneity, but the summary effect remained highly significant at each threshold. Population centile charts were also strongly associated with neonatal mortality, but generally showed a weaker association at all thresholds than absolute birth weight, because the summary ORs were closer to 1, especially compared to an absolute birth weight < 1.5kg or

2.0kg.The results for other measures, including ponderal index, birth weight < 2 standard deviations below the population mean, and fetal growth ratio (defined in the primary study as the observed birth weight over the population mean) varied (Figure 6.2).

The association between measures of being small for gestational age and neonatal morbidity are given in Figure 6.3. The analysis was subdivided into reported neurological morbidity (including seizures, HIE, IVH) and non-neurological morbidity (including hypoglycaemia, respiratory distress syndrome, cardiac failure) according to the definitions given in the primary studies. Birth weight <2.0kg was the most strongly associated with neurological morbidity, (OR 17.34, 95% CI 5.63 to 53.70) however this was based on a single study of 770 neonates. There was a significant association between weight below the 3rd, 5th and 10th centiles and neurological morbidity. Birth weight <10th centile according to customised growth chart and ponderal index ≤2.25 did not show a significant association with this outcome. For non-neurological morbidity, birth weight <3rd, 5th or 10th centiles on population chart and birth weight > 2 standard deviations (SD) below the population mean showed significant association with this outcome with odds ratios of a similar magnitude.

Infant outcomes

Figure 6.4 gives the association between measures of being small for gestational age and infant outcomes. Birth weight <1.5kg was strongly associated with infant morbidity on average (OR 21.57, 95% CI 6.31 to 73.70), however significant heterogeneity was present in the analysis. Birth weight <2.0kg and <2.5kg were also significantly associated with this outcome but the odds ratios were smaller. For

neurodevelopmental delay, birth weight <10th centile according to population chart was significantly associated with this outcome (OR 2.03, 95% CI 1.01 to 4.08)

Childhood and adolescent outcomes

A forest plot for the association of definitions of being small for gestational age with childhood and adolescent outcomes is given in Figure 6.5. Meta-analysis was performed to assess the association of birth weight <2.5kg with a composite group of adverse outcomes reported in primary studies (including obesity, hypertension, type 1 diabetes mellitus, asthma, hypercholesterolaemia, learning difficulties and strabismus). There was no significant association present (OR 0.98, 95% CI 0.87 to 1.10). A meta-analysis for birth weight <10th centile on population chart showed a small association that was just significant (OR 1.49, 95% CI 1.02 to 2.19) however there was significant heterogeneity present. When the analysis was restricted to learning difficulties or mental handicap, birth weight <3rd centile on population chart and <10th centile both showed a weak but significant association. There was no significant association between any birth weight standard and childhood obesity, hypertension, asthma, visual impairment or psychiatric diagnosis.

Adult outcomes

A forest plot of odds ratios for the association of measures of being small for gestational age and adult outcomes is given in Figure 6.6. A meta-analysis was performed for the association of birth weight <2.5 kg with a composite measure of adult morbidity (including obesity, hypertension, hypercholesterolaemia, type 2 diabetes mellitus, coronary heart disease, and polycystic ovarian syndrome). There was no significant association between birth weight <2.5kg, or birth weight <10th

centile on population chart, with this outcome. When individual morbidities were considered, birth weight <10th centile according to population chart was significantly associated with obesity according to a single study (OR 1.86, 95% CI 1.20 to 2.88). Birth weight <2.5kg showed a weak association with hypertension, diabetes mellitus or impaired glucose tolerance and cardiovascular mortality. Ponderal index (kg/m^3) <24 was also weakly associated with mortality from cardiovascular disease. Childhood or adulthood end stage renal disease showed a significant association with birth weight <10th centile on population chart.

Subgroup analyses

The results for subgroup analyses within the meta-analysis groups for each age group and birth weight standard are presented in Table 6.2. Few studies reported ethnicity in enough detail to permit subgroup analysis. For neonatal death, none of the subgroup analyses affected the magnitude or significance of the association between population chart <10th centile and this outcome. Limiting to a singleton population slightly weakened the association between birth weight <1.5kg and neonatal death, but did not affect birth weight <2.5kg for the same outcome. For childhood morbidity, singleton population, Caucasian population or year of birth ≥ 1990 did not significantly influence the results.

Birth weight as a continuous variable

There were seven papers that reported regression analysis using birth weight as a continuous outcome. These studies looked at adult hypertension (age 50 and 60 years) and hypercholesterolaemia, childhood obesity and hypertension, composite childhood metabolic risk index. Only one found a significant association (Anderrson

et al 2000): logistic regression analysis to examine the association between birth weight and hypertension (defined as treatment for hypertension and /or systolic BP ≥ 160mmHg and/or diastolic BP >95mmHg) found that at age 60, the OR was 0.96 (95% CI 0.92 to 0.99, p= 0.028 for change in risk of hypertension per 100g birth weight).

Direct comparison of absolute versus population centiles

Only two studies directly compared absolute birth weight and centile on population chart in the same population. For type 1 diabetes in childhood, birth weight <2.5kg had an OR of 0.68 (95% CI 0.22 to 2.12), and population chart <10th centile had an OR of 0.46 (95% CI 0.26 to 0.82) (Algert et al 2009). For neonatal mortality, birth weight <2.9kg had an OR of 2.64 (95% CI 1.45 to 4.82) and population chart <10th centile had an OR of 5.51 (95% CI 2.95 to 10.31) for the same outcome (Balcazar 1990).

Predictive ability of birth weight standards for neonatal death

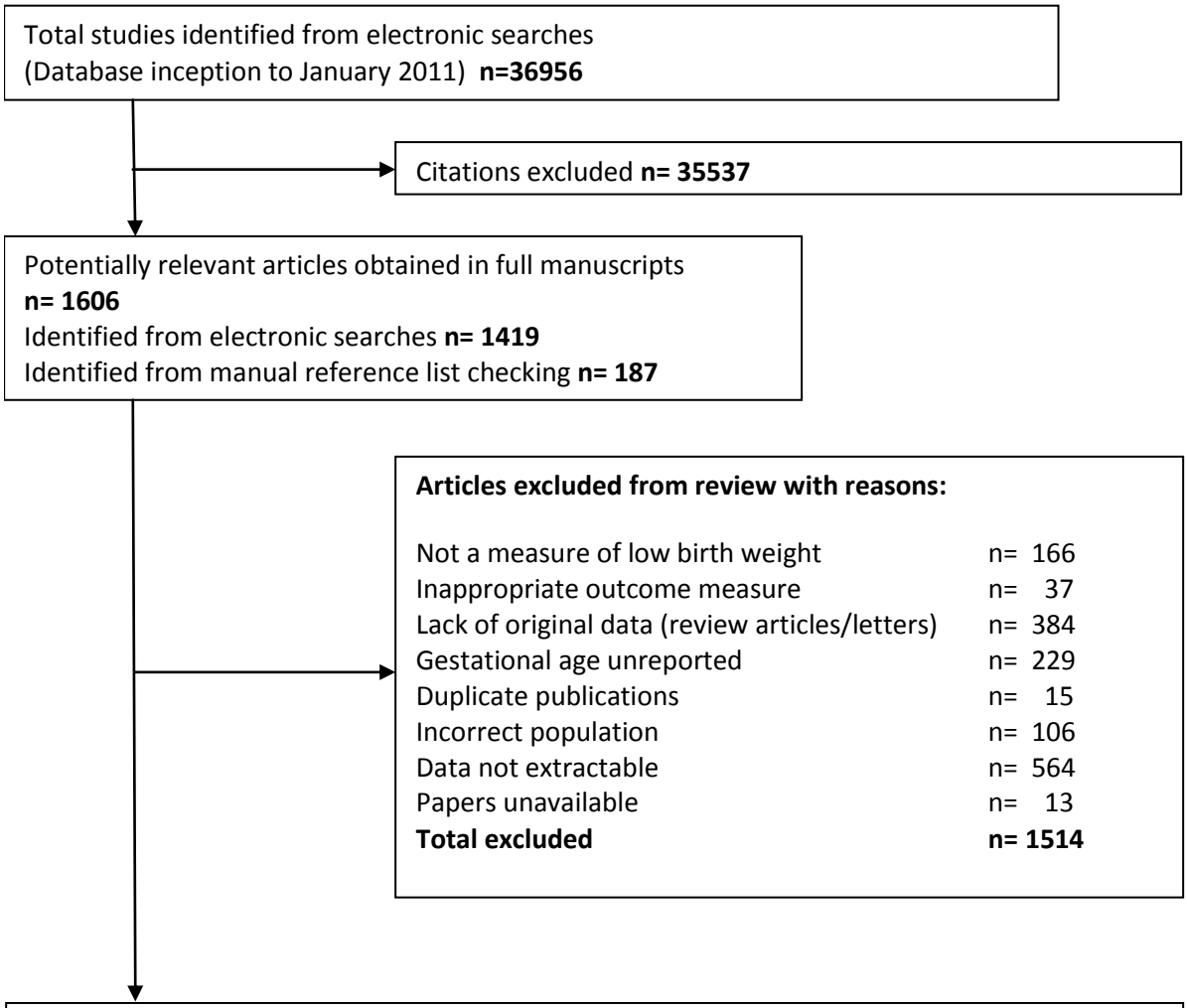
The outcome that had the strongest prognostic association overall with fetal growth restriction was neonatal death. For those birth weight tests with a large (OR>5) and statistically significant prognostic association for this outcome, their predictive ability for individual babies was summarised by using meta-analysis to calculate summary sensitivity, specificity and likelihood ratios (Table 6.3). These measures reveal the discriminative ability of each test and how test results modify a baby's odds of having a neonatal death. For each test the specificities and positive likelihood ratios were high, but the sensitivity and negative likelihood ratios were generally poor (Table 6.3). This can be explained by the fact that although a higher *proportion* of deaths

occurred within the low birth weight group, because this group represents a small fraction of the overall population, a large *absolute* number of deaths still occurred within the normal weight groups, and therefore sensitivity is low and the 'false negative' numbers are high, giving a poor negative likelihood ratio (close to 1). For example, the highest positive likelihood ratio was for 1.5kg, indicating that any baby under this weight multiplied their pre-test odds of neonatal death by 49.1 (95% CI: 27.3 to 88.5). However the negative likelihood ratio was only 1.01 (1.00 to 1.01), indicating that the odds of death barely change after a negative test result. Thus although being < 1.5kg substantially increases the odds of a poor outcome, being > 1.5kg does not increase the odds of a good outcome.

6.4.4 Publication bias for prognostic association

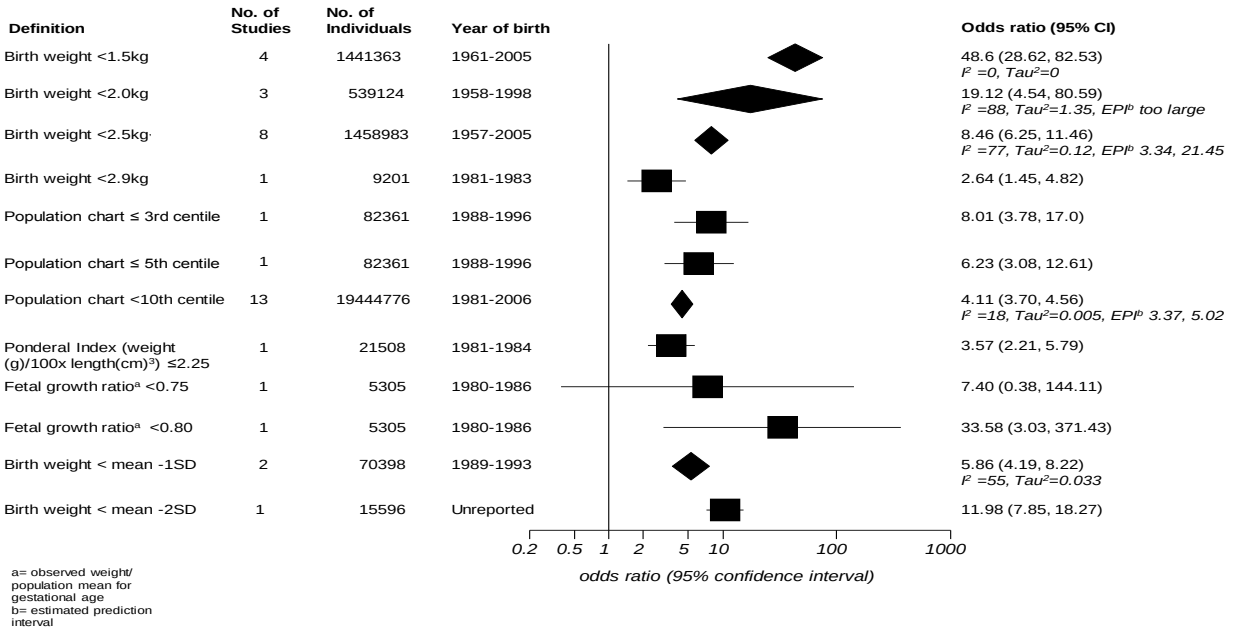
The Peters test was performed on all groups where there were 10 or more studies included in the meta-analysis (population <10th centile and neonatal mortality, birth weight <2.5kg and population <10th centile and childhood morbidity). There was no significant evidence of small study effects in any of the groups analysed (funnel plots not shown) (p values ranged from 0.326 to 0.996).

Figure 6.1 Study selection process for systematic review of the prognostic and predictive ability of current birth weight standards for short and long term outcomes



Selected for inclusion in systematic review	**n= 92**	**(23051541 individuals)**
Test	**No. of studies**	**No. of individuals**
Birth weight <2.5kg	31	2062710
Birth weight other thresholds	15	1468386
Population growth chart< 10th percentile	34	21944504
Population growth chart other percentiles	14	210506
Customised growth chart < 10th percentile	2	13359
Customised growth chart other percentiles	1	654
Ponderal Index	9	65885
Other measurements of fetal growth	10	87369

Figure 6.2 Forest plot of odds ratios for the association between birth weight standards and neonatal mortality. Single studies are represented by a filled box, and pooled studies (and 95% confidence intervals) by a filled diamond



a= observed weight/ population mean for gestational age
b= estimated prediction interval

Figure 6.3 Forest plot of odds ratios for the association between birth weight standards and neonatal morbidity. Single studies are represented by a filled box, and pooled studies (and 95% confidence intervals) by a filled diamond

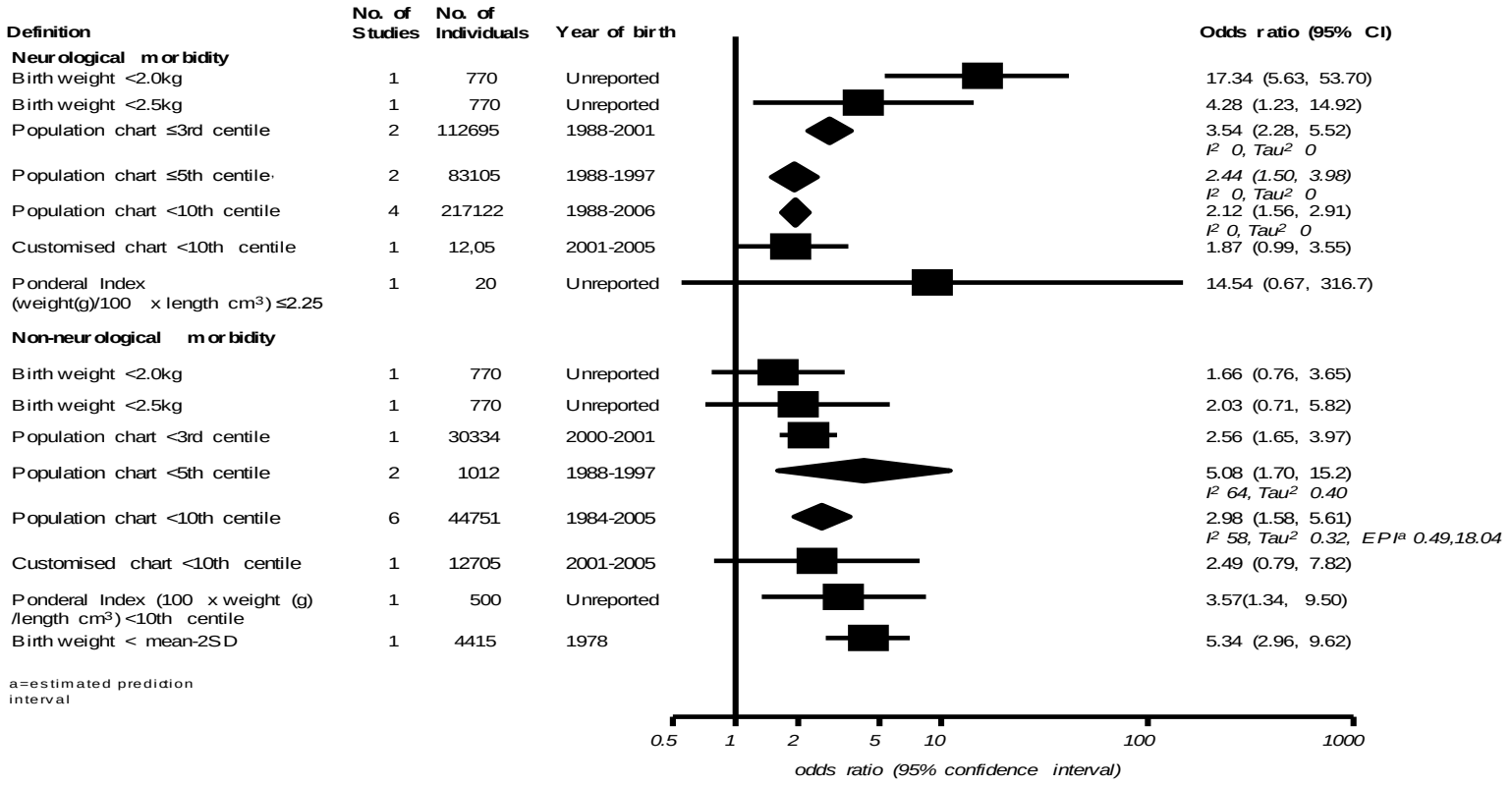


Figure 6.4 Forest plot of odds ratios for the association between birth weight standards and infant outcomes. Single studies are represented by a filled box, and pooled studies (and 95% confidence intervals) by a filled diamond

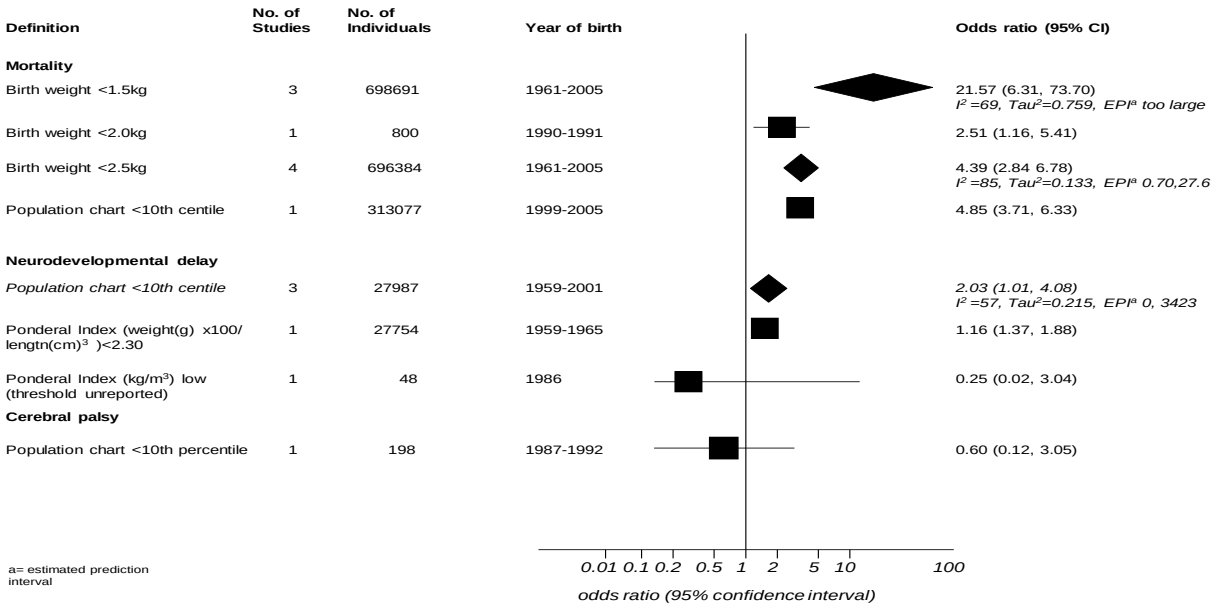


Figure 6.5 Forest plot of odds ratios for the association between birth weight standards and childhood outcomes. Single studies are represented by a filled box, and pooled studies (and 95% confidence intervals) by a filled diamond

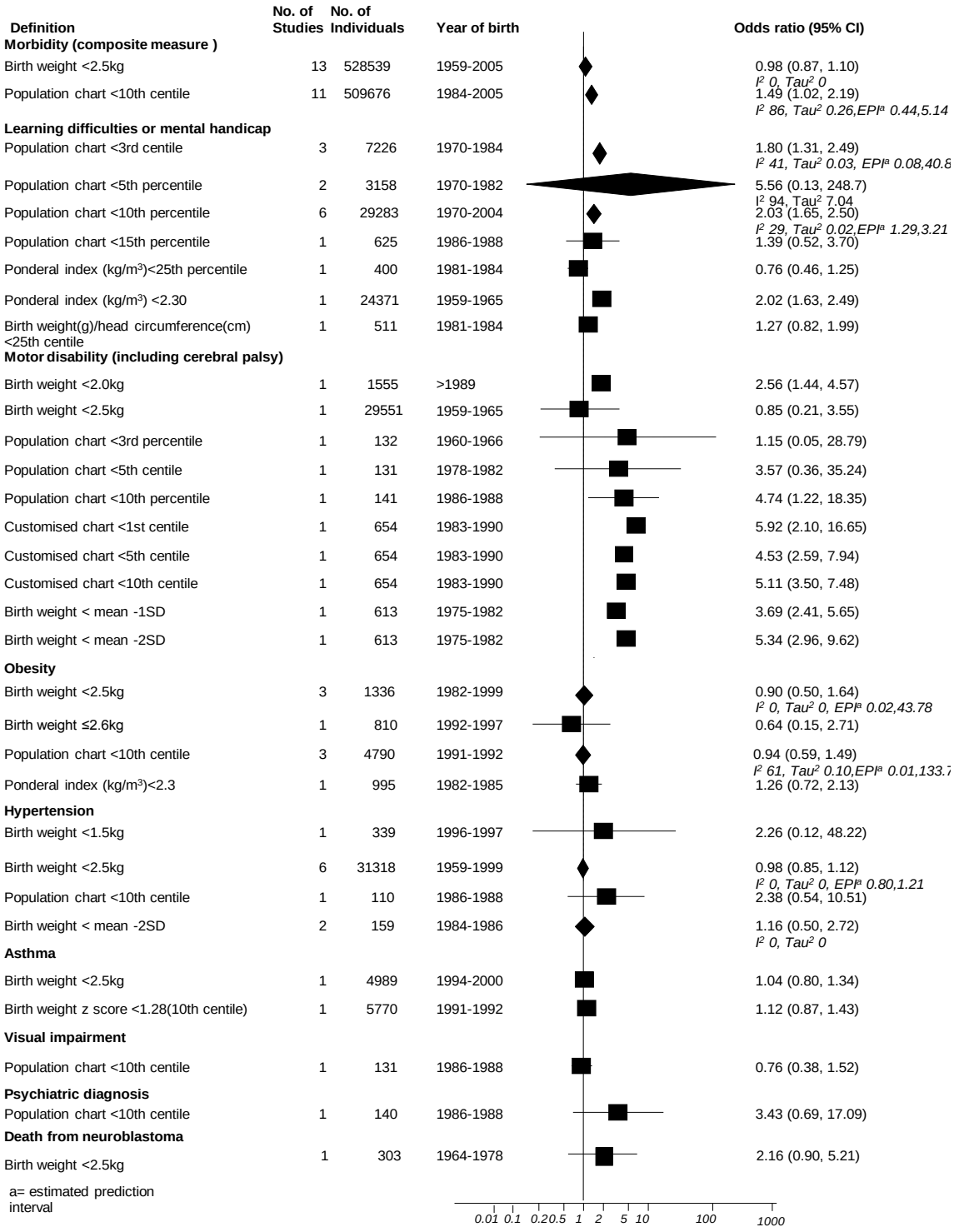
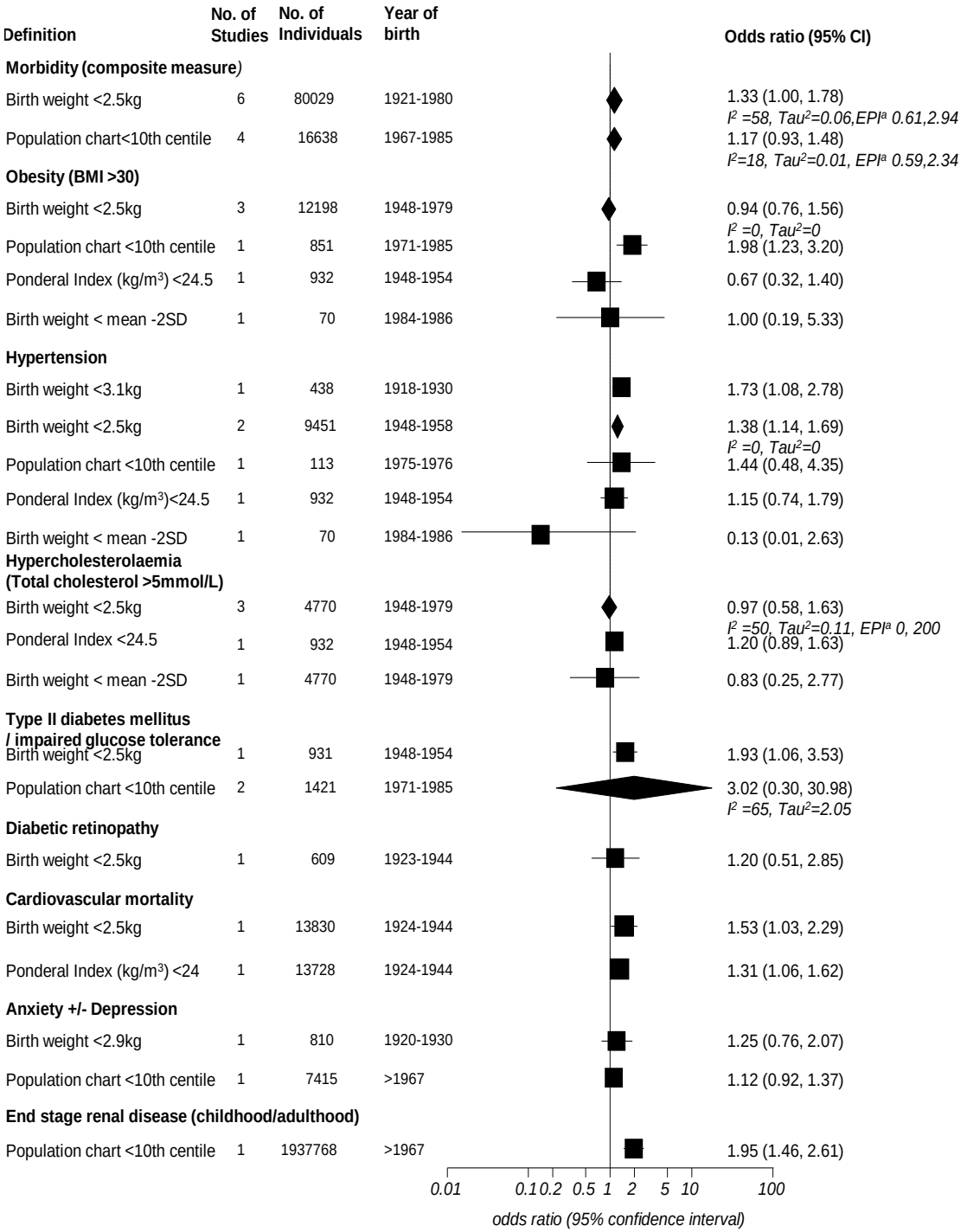


Figure 6.6 Forest plot of odds ratios for the association between birth weight standards and adult outcomes Single studies are represented by a filled box, and pooled studies (and 95% confidence intervals) by a filled diamond



a= estimated prediction interval

Table 6.1 Methodological quality of studies included in systematic review of birth weight standards for short and long term outcomes

	Number (%) of studies n= 92		
Quality Item	**Yes**	**No**	**Unclear**
Cohort study design	84 (91)	7 (8)	1 (1)
Population adequately described	90 (98)	0	2 (2)
Consecutive recruitment	42 (46)	12 (13)	38 (41)
Prospective recruitment	33 (36)	53 (58)	6 (6)
Appropriate outcome measure	92 (100)	0	0
Outcome measure blinded	10 (11.0)	1 (1)	81 (88)
>90% of individuals had outcome measure	42 (46)	41(44)	9 (10)
Index test and outcome measure described	36 (39)	5 (5)	51 (56)
Intervention between index test and outcome	5 (5)	0	87 (95)
Quality Classification			
High	56 (61)	-	-
Medium	27 (29)	-	-
Low	9 (10)	-	-

Table 6.2 Subgroup analyses according to birth weight standard and outcome, where possible, for study quality, ethnicity, year of birth of study population and singleton population

Birth weight standard	Number of studies	Subgroup	Odds ratio (95% CI)	Estimated prediction interval (EPI)	I^2 Tau^2
Neonatal death					
Birth weight <1.5kg	3	High quality studies	53.29 (30.08 to 94.39)	-	I^2=0, Tau^2=0
Birth weight <1.5kg	2	Singletons	41.85 (16.53 to 105.94)	-	I^2=0, Tau^2=0
Birth weight <2.5kg	4	Singletons	8.39 (4.90 to 14.36)	0.86 to 81.36	I^2=81, Tau^2=0.20
Birth weight <2.5kg	5	High quality studies	8.15 (5.76 to 11.54)	2.40 to 27.66	I^2=80, Tau^2=0.12
Birth weight <2.5kg	2	Year of birth ≥ 1990	9.74 (5.31 to 17.86)	-	I^2=91, Tau^2=0.17
Population chart <10th centile	6	Singletons	4.03 (3.88 to 4.18)	-	I^2=0, Tau^2=0
Population chart <10th centile	8	Year of birth ≥ 1990	4.23 (3.73 to 4.81)	3.23 to 5.55	I^2=31, Tau^2=0.01
Population chart <10th centile	4	Congenital anomalies excluded	4.01 (3.86 to 4.16)	-	I^2=0, Tau^2=0
Population chart <10th centile	6	Studies in USA/ Europe	4.04 (3.89 to 4.19)	-	I^2=0, Tau^2=0
Childhood morbidity					
Birth weight <2.5kg	5	Singleton	0.95 (0.63 to 1.44)	-	I^2=0, Tau^2=0
Birth weight <2.5kg	5	High quality studies	0.82 (0.63 to 1.07)	-	I^2=0, Tau^2=0

Birth weight <2.5kg	7	Ethnicity >90% Caucasian	0.99 (0.68 to 1.44)	-	I^2=0, Tau^2=0
Population chart <10th centile	4	Singleton	1.35 (0.82 to 2.24)	0.15 to 12.24	I^2=90, Tau^2=0.20
Population chart <10th centile	8	High quality studies	1.65 (0.96 to 2.83)	0.31 to 8.86	I^2=89, Tau^2=0.40
Population chart <10th centile	2	Year of birth ≥ 1990	0.67 (0.35 to 1.31)	-	I^2=76, Tau^2=0.19
Population chart <10th centile	2	Ethnicity Caucasian	3.70 (0.75 to 18.28)	-	I^2=91, Tau^2=1.22
Adult morbidity					
Birth weight <2.5kg	4	Singleton	1.41 (0.80 to 2.47)	0.14, 13.82	I^2=70, Tau^2=0.20
Birth weight <2.5kg	2	High quality studies	1.39 (1.14 to 1.69)	-	I^2=0, Tau^2=0

Table 6.3 Results for the predictive ability (sensitivity, specificity and likelihood ratios) of different birth weight standards for neonatal mortality

Birth weight standard	Sensitivity (95% CI)	Specificity (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
Birth weight <1.5kg (4 studies)	0.008 (0.004 to 0.146)	0.99 (0.99 to 1.00)	49.1 (27.3 to 88.5)	1.01 (1.00 to 1.01)
Birth weight <2.0kg (3 studies)	0.05 (0.03 to 0.07)	0.99 (0.99 to 1.00)	13.3 (2.27 to 78.28)	0.94 (0.85 to 1.02)
Birth weight <2.5kg (8 studies)	0.31 (0.19 to 0.47)	0.94 (0.88 to 0.97)	5.27 (3.57 to 7.76)	1.37 (1.15 to 1.62)
Population chart <3rd centile (1 study)	0.24 (0.12 to 0.41)	0.96 (0.96 to 0.96)	6.31 (3.57 to 11.14)	0.79 (0.66 to 0.94)
Fetal growth ratio <0.80 (1 study)	0.67 (0.09 to 0.99)	0.94 (0.93 to 0.95)	11.9 (3.87 to 32.52)	0.36 (0.07 to 1.75)
Birth weight < mean -2SD (1 study)	0.13 (0.09 to 0.19)	0.99 (0.99 to 0.99)	10.53 (7.25 to 15.28)	0.88 (0.83 to 0.92)

6.5 Discussion

Low birth weight showed a strong, consistent association with neonatal mortality. The relationship was highest at lower thresholds and gradually decreased (but remained strong) as the threshold increased. The absolute birth weight seemed to be more strongly related to this outcome than centiles on population weight charts, especially for thresholds of 1.5kg and 2.0kg. Restricting the analysis to singletons, year of birth since 1990, or by country of origin did not change the magnitude of the association. Other definitions of being small for gestational age were based on single studies and showed mixed results, but none appeared to be more strongly associated with neonatal

mortality than the absolute birth weight. A limitation, however, is that different numbers of studies contributed to each analysis, and there were few direct comparisons. Indeed, in the only study that compared absolute birth weight and centile chart in the same population, the association for birth weight below 10th centile was observed stronger than absolute birth weight <2.9kg for neonatal mortality. There was a lack of data in some analyses, e.g. customised centile charts and ponderal index in relation to adverse outcome, but as every effort was made to acquire both published and unpublished data nothing further could be done to address this. The results for neonatal morbidity were mixed, but no single definition of being small for gestational age appeared to be consistently more strongly associated with adverse outcomes than others. All of the birth weight and population chart thresholds assessed for predictive ability showed a high specificity and positive likelihood ratio for neonatal death, and thus babies who test positive are at a substantially higher risk of neonatal mortality. However, each test generally had a low sensitivity and negative likelihood ratio close to 1, and thus a negative test result does not improve the odds that a baby will not have a neonatal death.

For outcomes in childhood, there was a significant association between birth weight <10th centile according to population chart and a composite measure of morbidity. However when this analysis was restricted to a singleton population, this became non-significant. There was no significant association between any measure of low birth weight and childhood obesity, hypertension or asthma. Learning difficulties and motor disability were significantly associated with some reported definitions of being small for gestational age, but the associations overall were weak and no one measure showed a significantly stronger

relationship, including customised growth chart centiles. For adult outcomes, there was no consistent association seen between birth weight standards and adult health, although individual studies showed a weak association between birth weight <2.5kg and hypertension, cardiovascular mortality and diabetes.

The strength of this review lies in the methodology used. It complies with existing guidelines for the reporting of systematic reviews of diagnostic and observational studies[78;84-88] and uses the most up to date techniques for performing and interpreting meta-analysis.[102;130;131] An extensive literature search was performed in relevant databases with no language restrictions applied. Every effort was made to obtain the most complete dataset possible through contact with authors and experts in the field. Peters test showed that there was no evidence of small study bias.

There are several limitations to this review. Although every effort was made to control for potential confounding factors through subgroup analysis, due to the quality and reporting of the primary studies this was not always possible. The inclusion criteria were strictly limited to infants born at 37 weeks gestation or more, however the method of estimating gestation in the primary studies was often inaccurate. Very few studies used ultrasound measurement of crown-rump length at 10-13 weeks gestation, which is the most accurate method;[132] the majority used the mother's last menstrual period, some clinical examination of the newborn, which are less reliable and may have resulted in pre-term infants being included inadvertently.

Another limitation is that gestation may still influence outcome within the population of infants born at 37 weeks gestation or greater. It is known that

outcomes for babies at 37 weeks are worse than at 39 weeks. It is likely that the absolute birth weight thresholds captured more infants who were close to 37 weeks than the centiles did, and therefore this may account for the stronger association with adverse association seen in this group. The decision to restrict the review to infants born at term gestation meant that definitions for growth restriction at earlier gestations could not be assessed. This is clearly of importance due to the fact that if growth restriction is suspected antenatally these infants are often delivered prematurely, however it was felt that the potential for confounding in this group was also higher and that the objective should focus initially on tern born infants. Due to poor reporting in the primary studies, the potential to perform subgroup analysis according to ethnicity was limited. Although the results did not differ much when limited to a Caucasian population, it is known that Afro-Caribbean and Asian populations have smaller babies, and therefore it is likely that the same thresholds would not give the same results in all ethnic backgrounds.[133] Social class was not analysed as a subgroup, however previous epidemiological studies that have accounted for this have found that association between birth weight and cardiovascular risk factors persisted across social groups, suggesting that known and unknown confounding variables do not affect this relationship.[134]

Differences may exist between the birth weight standards that were combined within the analysis, particularly regarding the variety of different population charts used. It was not possible to compare different charts, however if each study used the appropriate chart for their population then this should not influence the results. Comparing different standards of birth weight through analyses using different populations may not give a true result. However, no

studies reported more than two standards in the same population, and only two studies compared absolute birth weight and population centile charts, limiting the ability to address this issue.

There is a vast literature exploring the relationship between low birth weight and adverse outcomes, using different methodologies to do so. The aim of this review was to consider the association and prediction of different thresholds of birth weight or centile charts, and therefore studies were excluded where 2 x 2 tables could not be obtained from the original paper or authors could not provide this. A complete assessment of the association of birth weight as a continuous variable with adverse health outcomes could not be made. In order to address this, a qualitative analysis of all the studies included in the review where regression analyses for the association of continuous birth weight with the health outcomes in question were reported. Only one study found a significant relationship between birth weight and adult blood pressure on regression analysis. Other systematic reviews performed in this field using birth weight as a continuous variable have shown mixed results. Owen et al examined the association between birth weight and blood cholesterol level, and found a weak association, however this analysis did not exclude pre-term infants. Huxley et al found an inverse association between birth weight and systolic blood pressure in children, adolescents, and adults, but again did not exclude pre-term infants from the analysis. Whincup et al found mixed results in the relationship between type II diabetes mellitus and birth weight. Nine out of 31 studies included in their systematic review showed a significant inverse relationship between birth weight and this outcome, again prematurity was not excluded.

Arcangeli et al recently conducted a systematic review examining the influence of small for gestational age (birth weight <10th centile or <2.5kg) and fetal growth restriction (birth weight <10th centile plus abnormal antenatal umbilical Doppler or placentation) on neurodevelopmental outcome in term infants. They found that small for gestational age infants had significantly lower neurodevelopmental indices than normally grown controls, but did not have sufficient data for precise results in the fetal growth restriction group. They acknowledge that some of the infants classified as small for gestational age may have had abnormal Dopplers, that were not reported in the primary studies.[135] This finding highlights the importance for improved reporting and individual patient data meta-analysis in this field.

The original literature published in support of the Barker hypothesis has been criticised for failing to control for potential confounding factors within their analysis.[52] Every effort was made to consider these within this analysis, and the findings with regard to childhood and adult health outcomes linked with the metabolic syndrome have been inconsistent. Where a composite outcome was used, no significant association for childhood or adult morbidity was seen. No significant association was present for childhood diabetes, hypertension or obesity. Weak associations were seen between birth weight and adult hypertension, diabetes and cardiovascular mortality, however meta-analysis was not possible for every outcome and therefore the results are based on the results of one or two studies.

Implications for clinical practice

Meta-analysis confirms that birth weight has a strong prognostic association with neonatal mortality, with low birth weight substantially increasing the risk of a poor outcome. However, although specificity and positive likelihood ratios were excellent, sensitivity was usually less than 0.5 and negative likelihood ratios were close to 1. This means that, compared to the pre-test risk of neonatal death (prevalence), babies with a low birth weight (test positive) are at a substantially increased risk, but the risk for those with a normal birth weight (test negative) does not change.

For example using the results for a threshold of 1.5kg (Table 6.3), for a population with a prevalence (risk) of neonatal death of 0.003 (3 in 1000), then following a positive test result a baby's risk of neonatal death becomes 0.13 (130 in 1000), but following a negative test result the risk remains at 0.003. The same pattern was seen for the other birth weight standards, and therefore being test negative does not reduce the odds of a poor outcome. None of the other health outcomes in the neonatal period or later life showed a strong enough prognostic association with measures of fetal growth restriction to warrant calculation of sensitivity, specificity and likelihood ratios.

Recommendations

Future research is necessary to identify a birth weight standard which can predict adverse health outcomes. Initially, it is important to compare the different standards across the same population to enable an unbiased comparison, and to further explore the standards which were less frequently reported and therefore could not be included in meta-analysis within this review, such as

ponderal index and customised centile charts. This could be performed through an individual patient data meta-analysis, where multiple definitions of fetal growth restriction could be compared across the same population, and factors such as ethnicity more adequately assessed.[121] An alternative option would be to perform further analysis on the large Scandinavian birth registries, which record a variety of birth anthropometry that can be linked to health outcomes.[136]

Finally, it is likely that more accurate predictions could be made using birth weight as a continuous variable, rather than dichotomising it using a threshold as is currently the general practice.[137] Using the raw birth weight, and potentially considering non-linear associations between weight and outcome risk, would allow outcome risk to be estimated for each individual accordingly to their specific weight value. This would increase the power and is likely to improve predictive accuracy. At the meta-analysis level, individual patient data are required to undertake such an analysis approach, which was beyond the scope of our review. [138] Even when analysed on its continuous scale, birth weight on its own may not have sufficiently accurate predictive ability, and so then birth weight in combination with other factors should be explored to predict adverse outcome in clinical practice. This could be achieved using a prognostic model.[139]

6.6 Conclusion

Birth weight tests are strongly associated with neonatal mortality and morbidity, especially at lower absolute birth weight thresholds, and test positive babies (small for gestational age) are at a substantially increased risk of neonatal mortality. However, babies who test negative do not have a decreased risk of a

neonatal mortality. The association between low birth weight and childhood and adult morbidity was inconsistent. Further research is required to identify the optimum definition of being small for gestational age that helps best predict the risk of adverse outcomes, and this may require using birth weight as a continuous variable, developing prognostic models also containing other factors, and using individual patient data meta-analysis.

CHAPTER 7: SYSTEMATIC REVIEW AND META – ANALYSIS OF THE PROGNOSTIC AND PREDICTIVE ABILITY OF APGAR SCORE AT BIRTH FOR SHORT AND LONG TERM OUTCOMES

7.1 Abstract

7.1.1 Background

The purpose of this systematic review was to evaluate the relationship between Apgar score and subsequent adverse outcome and to explore the predictive ability of this test, including the optimal threshold to define a low score.

7.1.2 Methods

Systematic review of the literature, with random effects meta-analysis, to compute summary odds ratios (OR) to assess prognostic association, and summary sensitivity, specificity and likelihood ratios, to assess predictive ability.

Electronic searches were performed from database inception until June 2011 without language restrictions. The reference lists of selected articles were screened and

authors contacted. Studies were selected by 2 reviewers if a low Apgar score, by any threshold, was related to neonatal or long term outcomes.

7.1.3 Results

87 manuscripts were selected for final inclusion, with a total of 3690080 neonates. Meta-analysis performed according to threshold of Apgar score, showed that a low Apgar score was strongly associated with neonatal mortality, particularly in a population born at term (≥ 37 weeks gestation) or with normal birth weight (≥ 2.5kg) The largest associations in this population were seen with a 5 minute Apgar score ≤ 1 (single study, OR 2209.16, 95% CI 425.88 to 11000) and a 10 minute Apgar score ≤ 3 (single study, OR 1417.75, 95% CI 915.99 to 2194.36). Raising the Apgar score at a particular time reduced the strength of association, which for the 1 minute Apgar score became non-significant at a threshold of ≤9. In a pre-term population, the magnitude of the association with neonatal mortality was smaller, with the strongest association being at a 10 minute score ≤ 3 (single study, OR 66.49, 95% CI 45.00 to 98.22). For neonatal morbidity, significant association was seen at a number of thresholds across all of the populations examined, but the magnitude of the association was smaller and the prediction intervals crossed 1. In a term population, there was a significant association between a low Apgar score and cerebral palsy at all thresholds examined, with the largest association seen at a 5 minute Apgar score ≤ 3 (3 studies, OR 46.35, 95% CI 11.21 to 191.59). When the predictive ability of the Apgar score was considered, the specificity and positive likelihood ratios were generally large at low scores, however, the corresponding sensitivity and negative likelihood ratios were poor.

7.1.4 Conclusion

A low Apgar score at birth is strongly associated with neonatal mortality, morbidity and childhood cerebral palsy, particularly in a population born at term, or with normal birth weight. Further research is required to identify the subgroups in which Apgar score may best predict adverse outcome. This may be done through individual patient data meta-analysis. Prognostic models also containing other factors may optimise the prediction of adverse outcome on an individual level.

7.2 Introduction

Virginia Apgar score initially introduced her score in 1953 with the aim of focusing attention on the newborn infant and to identify those in need of resuscitation.[62] It is described in Table 2.3. The score is usually recorded at 1, 5 and 10 minutes following delivery.[61]The total score may be reduced by any factor that causes compromise to the neonate, and the validity of using the results to predict adverse outcomes has been questioned.[81] Conventional thresholds defined on the basis of the original cohort used to define the Apgar score, and subsequent biochemical studies, grouped a score of 0 to 3 at 1 minute as ‘poor’, 4 to 6 ‘fair’ and 7 to 10 as ‘good’.[140] This classification was ascribed 50 years ago, since which time there have been significant changes in neonatal care and a large body of work examining the relationship between a low Apgar score and adverse outcome. However, there remains a lack of consensus regarding the threshold that defines a ‘low’ score, and the timing of measurement to which the greatest importance can be attached.[82] The American Academy of Pediatrics recommends that it is inappropriate to use a low

Apgar score alone to diagnose asphyxia, or predict neurological adverse outcomes in the neonate.[61]

The aim of this review was to examine the association between a low Apgar score and adverse outcomes, considering subgroups according to gestational age, and to assess the predictive ability and optimum threshold of Apgar score for adverse outcome.

7.3 Methods

The methods employed are those outlined in Chapter 4, with those specific to this review detailed below.

7.3.1 Data sources and searches

Electronic searches were performed with the aim of capturing neonates with Apgar score recorded at birth (at either 1, 5 or 10 minutes of age). Searches were performed by the author from database inception until June 2011. The search strategy employed in Medline is given in Appendix 13. This was adapted for use in other electronic databases.

7.3.2 Study selection

Studies were selected if they examined the association between Apgar score at birth weight and an adverse outcome in live born neonates.

7.3.3 Data extraction and quality assessment

The data extraction tool employed in this review is given in Appendix 6. A low Apgar score recorded at 1, 5 and 10 minutes, defined by any threshold was included (e.g.5

minute Apgar score ≤ 6). Where multiple thresholds or outcome measures were reported, 2 x 2 tables were created for each individual threshold and outcome. Care was taken not to include each individual more than once in any given meta- analysis. Where possible, if a study reported data for pre-term (born at <37 weeks gestation) or low birth weight (< 2.5kg) infants separately to that of term born infants, a 2 x 2 table was created for each subgroup to enable further analysis. Data were extracted by the author, and partly in duplicate by a second reviewer (see acknowledgements).

All studies were assessed fully using the STARD and QUADAS checklists[90;91] (Appendix 2 and 3). The elements felt to be most relevant to systematic reviews assessing prognostic association and predictive ability, as described in section 4.4, were used to assess the overall quality of the included studies. A study meeting four or more of the criteria was considered to be of high quality, three moderate quality, and 2 or less of low quality.

7.3.4 Data synthesis

The 2 x 2 tables were used to compute odds ratios (OR) and 95 % confidence intervals (CI) for each index test-outcome pair, and pooled the results for each index test (considering each definition and threshold of growth as a separate test) using meta-analysis.

Summary OR data is presented in forest plots. Meta-analyses were performed where 2 or more studies reported the same index test and outcome measure. The primary outcomes were considered to be neonatal mortality and a composite measure of morbidity (up to 28 days), infant mortality (up to 1 year), cerebral palsy and a composite measure of other childhood morbidity. For each outcome group, separate

analysis was performed for a preterm (<37 weeks gestation) or low birth weight (<2.5kg) population; term born (≥ 37 weeks gestation) or normal birth weight (≥ 2.5kg) neonates; and an unrestricted population in terms of gestational age or birth weight. Predictive ability was assessed by calculating specificity, sensitivity and likelihood ratios, for the outcomes of neonatal mortality and cerebral palsy (i.e. non-composite measures) where the OR was greater than 5.

When a composite outcome measure was used, attempts were made to select the most consistent threshold and outcome across the analysis: for a 1 minute Apgar score ≤ 3 and neonatal morbidity in a pre-term population, IVH and other cerebral abnormalities were the most commonly reported outcomes therefore were selected preferentially to non-neurological morbidity. Subgroup analysis according to individual conditions within the morbidity groups was also performed where possible.

Within the largest meta-analyses subgroup analysis was performed to examine the effect of potential confounding factors. Study quality, exclusion of congenital anomalies, birth of the study population during or after 1990 (due to recent advances in antenatal and neonatal care), and location of study (USA, Europe, Australia and New Zealand versus elsewhere) were considered to be important factors which may influence the strength of the association between Apgar score and adverse outcome.

7.4 Results

7.4.1 Literature identification and study characteristics

As described in Figure 7.1, after an initial review of 4692 citations derived from electronic searches, 275 studies were selected for full review, with a further 73 identified from manual reference checking of included studies and review articles, or

hand searching of major journals. 87 manuscripts were selected for final inclusion, with a total of 3690080 neonates. Four of these studies were included following provision of further data from the authors.

The characteristics of the included studies are given in Appendix 14. The reference list of included studies is given in Appendix 15. A number of studies contained duplicate populations with each other. If there was no difference in the population, index test threshold or outcome measure then the least complete study was excluded, but where different index test thresholds, subgroups or outcome measures were reported care was taken to include each individual only once in any meta-analysis and in the overall numbers included in the review. The majority of studies reported the 1 and/or 5 minute Apgar scores and the most commonly used thresholds to define a low Apgar score were ≤ 3 and ≤ 6. For neonatal mortality, data were available for all thresholds of the 1 minute Apgar score in all three populations, and for the 5 minute Apgar score in a term population. The thresholds reported for other outcomes varied.

7.4.2 Study quality assessment

The results for the quality assessment of included studies are given in Table 7.1. The majority of studies were of cohort design (80%). 56% of included studies were classified as being high quality according to the pre-specified criteria given in section 4.4. Most studies described the population adequately, and all had an appropriate outcome measure. However, descriptions of the index test and outcome measures were poor, with only 10% of included studies reporting these in a reproducible

fashion. This particularly applied to the Apgar score, with very few studies describing the personnel or technique employed.

Where possible, subgroup analysis was performed using high quality studies only, and the results are presented in Table 7.2.

7.4.3 Data analysis

Prognostic association between a low Apgar score and neonatal mortality

A forest plot of the prognostic association of the 1, 5 and 10 minute Apgar score, at a variety of thresholds, for an unrestricted population, is given in Figure 7.2. There was a significant association between a low Apgar score and mortality at all thresholds examined, with the strongest being a 10 minute Apgar score ≤ 3 (OR 616.72, 95% CI 461.52 to 824.10). This was based on a single study. A 10 minute Apgar score ≤ 6 (single study, OR 392.08, 95% CI 304.70 to 504.52) and a 5 minute score of ≤ 3 (4 studies, OR 290.88, 95% CI 183.91 to 460.07) also showed a very strong association with neonatal mortality in this population. In general, the higher the 1 minute Apgar score the weaker the association with neonatal mortality, and a low 5 and 10 minute Apgar score were more strongly associated with mortality than a low 1 minute score. The data for most thresholds at 1 minute were based on a single study (Apgar and James 1962). At thresholds where meta-analysis was possible there was significant heterogeneity present.

As reported in Figure 7.3, when limited to a population defined as either born at term (≥ 37 weeks) or with a birth weight ≥ 2.5 kg, the association was again significant at all thresholds, and appeared stronger than in an unrestricted population. The largest associations were seen with a 5 minute Apgar score ≤ 1 (single study, OR 2209.16,

95% CI 425.88 to 11000) and a 10 minute Apgar score ≤ 3 (single study, OR 1417.75, 95% CI 915.99 to 2194.36). Again there was significant heterogeneity present in the majority of meta-analysis groups.

When the population was limited to infants either born pre-term (< 37 weeks gestation) or low birth weight (<2.5kg), a low Apgar score by any threshold was still significantly associated with neonatal mortality (Figure 7.4), and the association was strong at most thresholds. However, the odds ratios were smaller than in the term/ normal birth weight population.

Prognostic association between a low Apgar score and neonatal morbidity

The association of the Apgar score with a composite measure of neonatal morbidity is presented in Figure 7.5. In an unrestricted population, there was a significant association between a low Apgar score by any threshold and neonatal morbidity. The strongest association was seen at a 5 minute Apgar score of ≤ 4 (single study, OR 25.26, 95% CI 10.14 to 62.91). At the thresholds where meta-analysis was possible, the strength of association was similar: 1 minute ≤ 3 had OR 3.32 (95% CI 1.80 to 6.17), 1 minute ≤ 6 OR 2.44 (95% CI 1.37 to 4.33) and 5 minute ≤ 6 OR 3.08 (95% CI 1.55 to 6.12). However, in all of these the EPI had a lower threshold below 1, giving uncertainty whether the association would be significant if tested in a new study.

When limited to a term/ normal birth weight population (Figure 7.5), the association between a low Apgar score and neonatal morbidity was again significant at all thresholds, with the strongest association at a 10 minute Apgar score ≤ 4, however this was based on a single study and the confidence interval was very large (OR 64.22, 95% CI 3.98 to 1037.39). It was only possible to calculate a prediction interval

for a 1 minute Apgar score ≤ 3 and neonatal morbidity in this population, and this had a lower threshold below 1 and was very broad (OR 11.83, 95% CI 3.36 to 41.65, EPI 0.06 to 2409).

For the pre-term/ low birth weight population, a low 10 minute Apgar score was not significantly associated with neonatal morbidity. A 1 minute Apgar score ≤ 3 and 5 minute Apgar score ≤6 both showed a significant association with morbidity through meta-analysis. It was only possible to calculate a prediction interval for 1 minute Apgar score ≤ 3 in this population and this had a lower threshold below 1 (5 studies, OR 2.36, 95% CI 1.32 to 4.22, EPI 0.31 to 17.73).

Subgroup analysis was performed where possible to investigate the association of Apgar score with individual morbidities rather than the composite outcome. This was only possible in a the pre-term population, where the association of a low Apgar score, defined as ≤ 3 at 1 minute, and neonatal IVH was calculated. This showed a significant association (OR 3.22, 95% CI 2.50 to 4.14, I^2= 12.4%, Tau^2 = 0.01; EPI 1.55 to 6.68) based on 4 studies.

Prognostic association between a low Apgar score and infant mortality

The association between a low Apgar score and infant mortality (defined as death up to 12 months of age) is given in Figure 7.6. The prognostic association for this outcome was significant at all thresholds of the Apgar score in both a pre-term/ low birth weight and term/ normal birth weight populations, with the largest association seen in the term population at a 5 minute Apgar score of ≤ 3 (OR 36.40, 95% CI 5.84 to 226.80). However, there was significant heterogeneity present in all of the meta-analyses possible, and the prediction intervals all crossed 1.

Prognostic association between a low Apgar score and cerebral palsy

As presented in Figure 7.7, when an unrestricted population was considered, a low Apgar score by any threshold showed a significant prognostic association with childhood cerebral palsy. The largest association was seen at a 5 minute Apgar score ≤ 3 (2 studies, OR 39.31, 95% CI 8.64 to 178.75). A strong association was also present at a 5 minute Apgar score ≤ 6 (2 studies OR 17.38, 95% CI 5.01 to 60.29), however these were the only thresholds where meta-analysis was performed and significant heterogeneity was present in both cases.

When limited to studies reporting neonates born at term or with normal birth weight, the association was significant at all thresholds, with the strongest association again seen at a 5 minute Apgar score ≤ 3 (3 studies, OR 46.35, 95% CI 11.21 to 191.59).

However, there was significant heterogeneity present in the analysis and the prediction interval was very large, indicating substantial uncertainty regarding the true magnitude of association.

In the pre-term/ low birth weight population, the association between low Apgar score and cerebral palsy was significant at all thresholds of the 1 minute Apgar score, and was largest at a 5 minute Apgar score of ≤ 3 (3 studies, OR 8.18, 95% CI 4.83 to 13.84). There was no significant heterogeneity within the analysis but the prediction interval crossed 1. A 5 minute Apgar score ≤ 4 or 5, and a 10 minute Apgar score ≤ 3 were not significantly associated with cerebral palsy, although these results were based on relatively small numbers of individuals.

Prognostic association between a low Apgar score and other childhood morbidity

The association between a low Apgar score and childhood morbidity (other than cerebral palsy) is presented in Figure 7.8. In an unrestricted population, there was a small but significant association between a low 1 minute Apgar score and the composite measure of childhood morbidity (≤3 3 studies, OR 2.14, 95% CI 1.46 to 3.15; ≤ 6 3 studies, OR 2.92, 95% CI 1.51 to 5.67). The prediction intervals crossed 1 in both cases. At a threshold of ≤ 7, the association was non-significant. A 10 minute score ≤9 had the largest association with morbidity in this population, however this was based on a single study (OR 9.95, 95% CI 3.68 to 35.65). When limited to a term/ normal birth weight population, the association was significant at a 1 minute Apgar score ≤ 3, but this was based on a single study and meta-analysis of other thresholds of the 1 minute Apgar score were non- significant. A 5 minute Apgar score of ≤3 and ≤6 both showed a significant association of a similar magnitude (5 studies, OR 3.11, 95% CI 1.40 to 6.92 and 6 studies, OR 3.53, 95% CI 1.74 to 7.13 respectively). However significant heterogeneity was present in both groups.

For a preterm/ low birth weight population, the 1 minute Apgar score was not associated with morbidity at any threshold, but the 5 minute Apgar score showed a significant association at both thresholds examined (≤3 5 studies, OR 4.11, 95% CI 1.25 to 13.54 and ≤ 6 4 studies, OR 2.01, 95% CI 1.03 to 3.91). Again, the prediction intervals crossed 1 in both cases.

Subgroup analysis according to individual conditions was possible in three groups. For a pre-term population, where a 5 minute Apgar score ≤ 3 was considered, there

was no significant prognostic association for neurologic disability (other than cerebral palsy, by definitions used in the primary studies) (3 studies OR 2.23, 95% CI 0.73 to 6.82, I^2=50%, Tau^2=0.49, EPI 0 to 221765). In a term population for the same outcome and Apgar score, the association was significant (3 studies, OR 2.12, 95% CI 1.15 to 3.92, I^2= 55%, Tau^2= 0.165, EPI 0 to 1438.9). At a 5 minute Apgar threshold of ≤ 6, there was a significant association with childhood epilepsy in a term/ normal birth weight population, as determined by meta-analysis of 2 studies (OR 3.84, 95% CI 2.47 to 5.96, I^2=0, Tau^2=0).

Subgroup analyses

The results for subgroup analyses for factors considered to be potential sources of heterogeneity are presented in Table 7.2. In a pre-term/ low birth weight population, limiting the analysis to 2 studies with populations born in 1990, or later, made the association between 1 minute Apgar score ≤ 3 and neonatal mortality non-significant. None of the other factors considered, including where possible high quality studies, exclusion of neonates with congenital anomalies, and location of study changed the significance of the association for this population and outcome.

In a term/ normal birth weight population, none of the subgroup analyses changed the significance of the association with mortality at any threshold, but limiting to high quality studies did eliminate heterogeneity from the analysis at 1 minute Apgar score thresholds of ≤ 3 and ≤ 6, and increased the magnitude of the association in both cases (OR 104.35, 95% CI 59.73 to 102.31 and OR 61.86, 95% CI 36.11 to 105.96 respectively).

In an unrestricted population, again subgroup analysis did not change the significance of the association of a low Apgar score at any threshold with neonatal mortality, or eliminate significant heterogeneity, with the exception of limiting to high quality studies in the analysis for a 1 minute Apgar score ≤ 3. In this case the heterogeneity was eliminated but the strength of the association was reduced and the prediction interval crossed 1 (3 studies OR 16.91, 95% CI 9.52 to 30.06).

For childhood morbidity, limiting the analysis to studies performed in Europe, the USA or Australia or New Zealand made the association between a 5 minute Apgar score ≤ 3 and this outcome non-significant in a pre-term population. For a term/normal birth weight population, limiting the analysis to studies where neonates with congenital anomalies were excluded eliminated heterogeneity in the analysis at a 5 minute Apgar score of ≤ 6, and the prediction interval also became significant (5 studies, OR 4.38, 95% CI 3.19 to 6.01, EPI 2.62, 7.33). However, the same subgroup analysis at the same threshold in an unrestricted population made the association non-significant.

For neonatal morbidity, limiting the analysis to high quality studies, year of birth ≥ 1990, or those excluding congenital anomalies eliminated heterogeneity within the analysis for 1 minute Apgar score ≥ 3 in a pre-term/ low birth weight population, without affecting the magnitude of the association. No subgroup analysis was possible for this outcome in a term or unrestricted population.

Subgroup analysis for the outcome of infant mortality was only possible in a pre-term population, and none of the analyses performed significantly affected the results or the magnitude of heterogeneity present.

Predictive ability of the Apgar score for adverse outcomes

The sensitivity, specificity and likelihood ratios of the Apgar score for neonatal mortality are presented in Table 7.3, and for cerebral palsy in Table 7.4. In general, the Apgar score at low thresholds had a high specificity and positive likelihood ratio for both outcomes. However, the associated sensitivity and negative likelihood ratios were low. For example, the highest positive likelihood ratio was for a 10 min score ≤ 3, indicating that any baby with a score less than this threshold multiplied their pre-test odds of neonatal death by 958.99 (95% CI: 698.99 to 1315.9). Assuming a pre-test probability of 0.003 of neonatal death,[3] having a score below this threshold increases the probability of death to 0.74. However the negative likelihood ratio was only 0.68 (0.59 to 0.77), giving a post-test probability of 0.002, therefore having an Apgar score of ≥ 3 at 10 minutes does not substantially change the risk of neonatal death. The formulae for these calculations are given in Appendix 16.

Publication bias for prognostic association

There was only one meta-analysis group large enough to perform the Peters test. For a pre-term/low birth weight population, 10 studies were included in the analysis for an Apgar score of 5 minutes ≤ 6, and neonatal mortality. The Peters test suggested a significant small study effect (p= 0.016

Figure 7.1. Study selection process for systematic review of the prognostic and predictive ability of Apgar score for short and long term outcomes

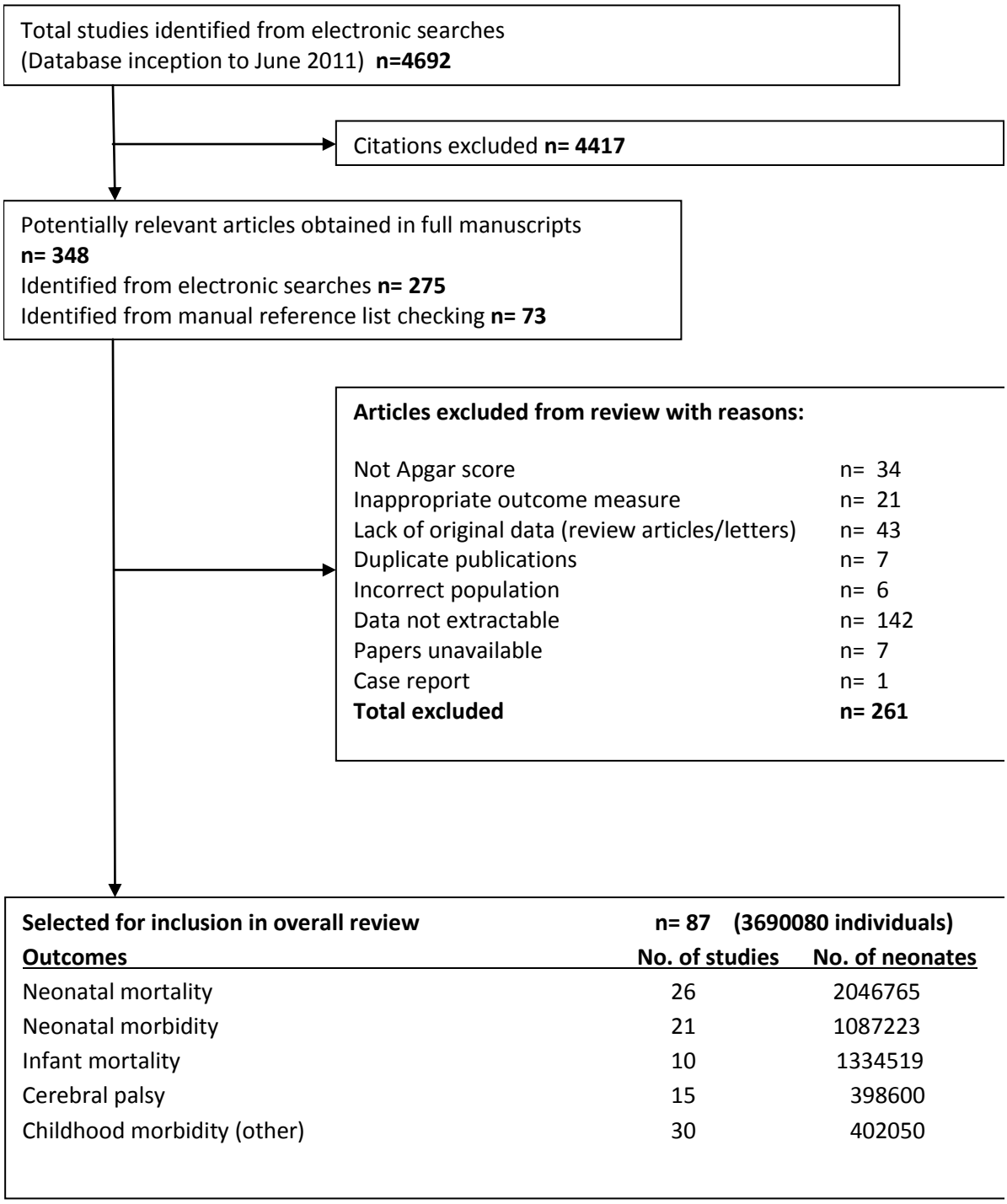
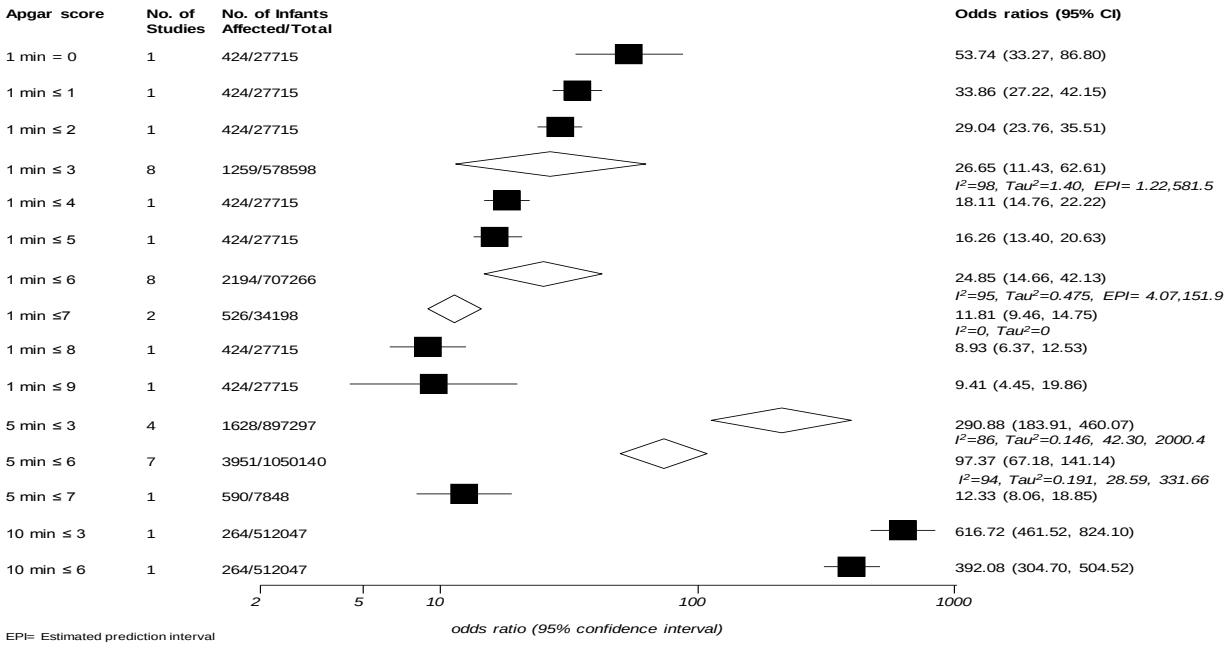
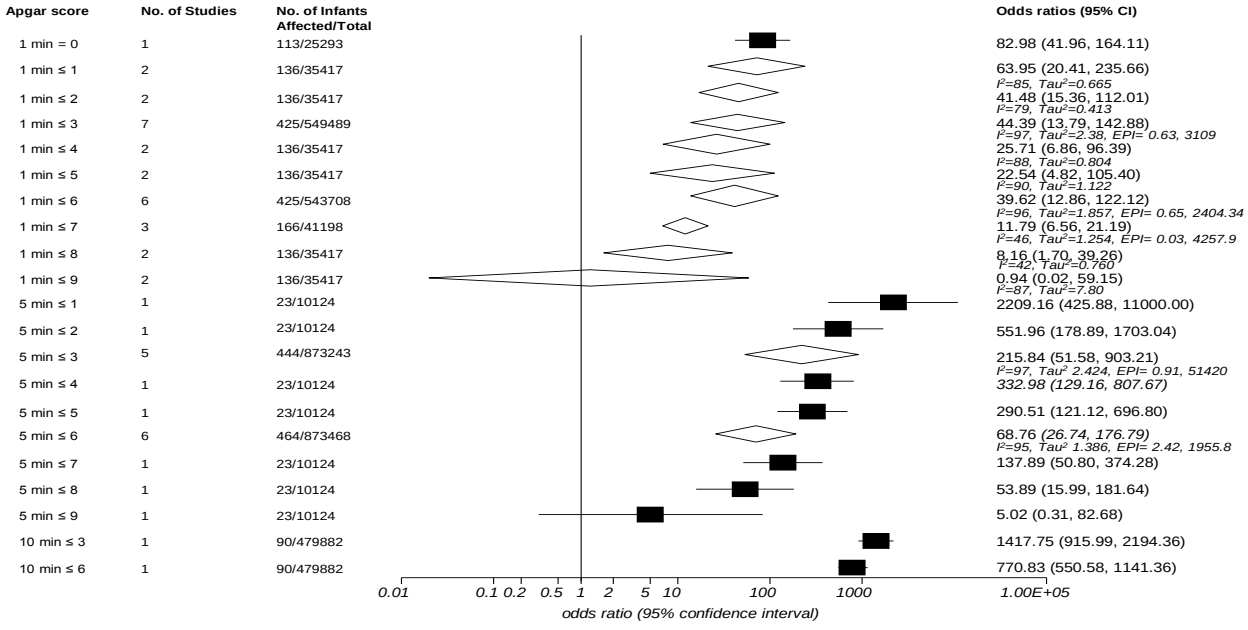


Figure 7.2 Forest plot of odds ratios for the association between Apgar score and neonatal mortality in an unrestricted population. Single studies are represented by a filled box, and pooled studies (and 95% confidence intervals) by a diamond



EPI= Estimated prediction interval

Figure 7.3. Forest plot of odds ratios for the association between Apgar score and neonatal mortality in a term/ normal birth weight population. Single studies are represented by a filled box, and pooled studies (and 95% confidence intervals) by a diamond



EPI= Estimated prediction interval

Figure 7.4 Forest plot of odds ratios for the association between Apgar score and neonatal mortality in a preterm/low birth weight population. Single studies are represented by a filled box, and pooled studies (and 95% confidence intervals) by a diamond

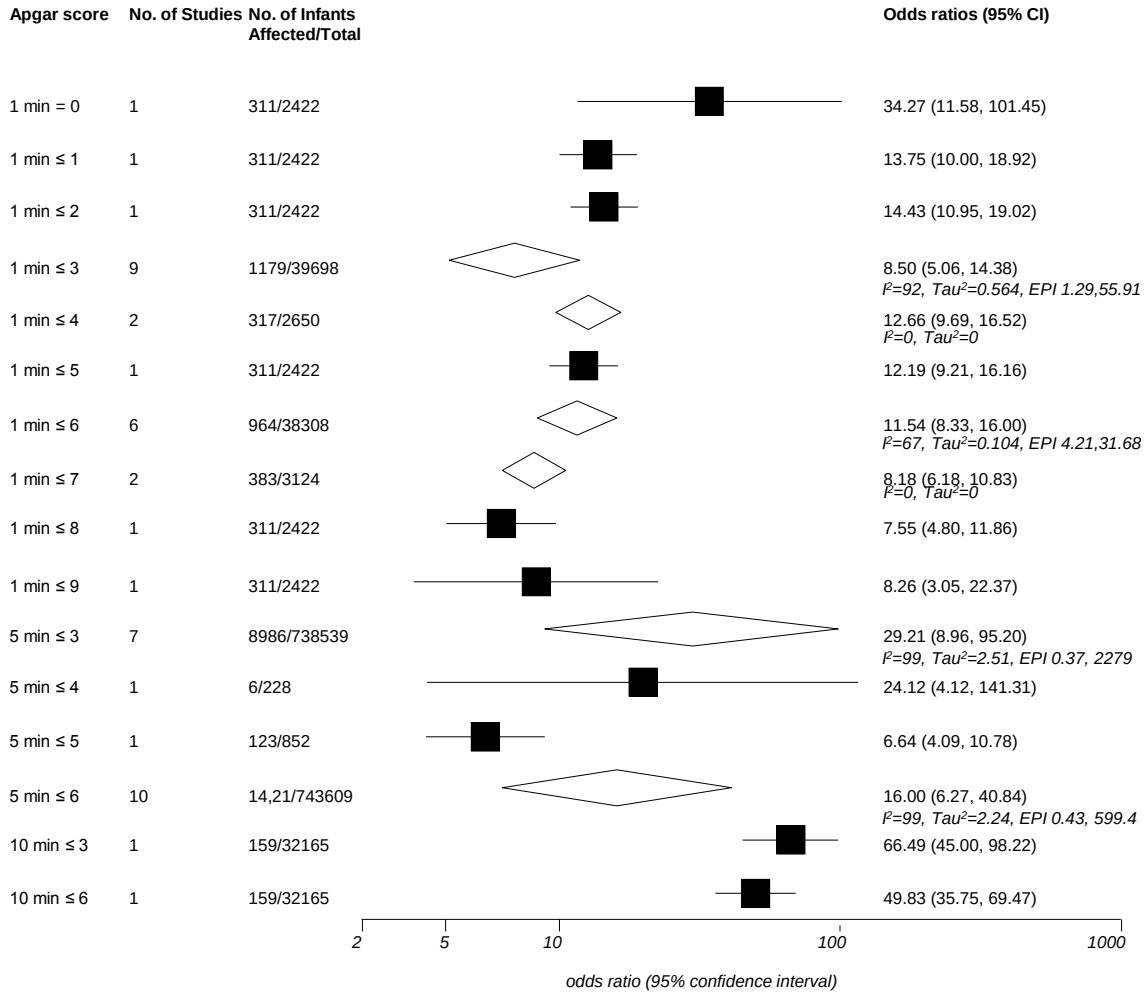


Figure 7.5 Forest plot of odds ratios for the association between Apgar score and neonatal morbidity. Single studies are represented by a filled box, and pooled studies (and 95% confidence intervals) by a diamond

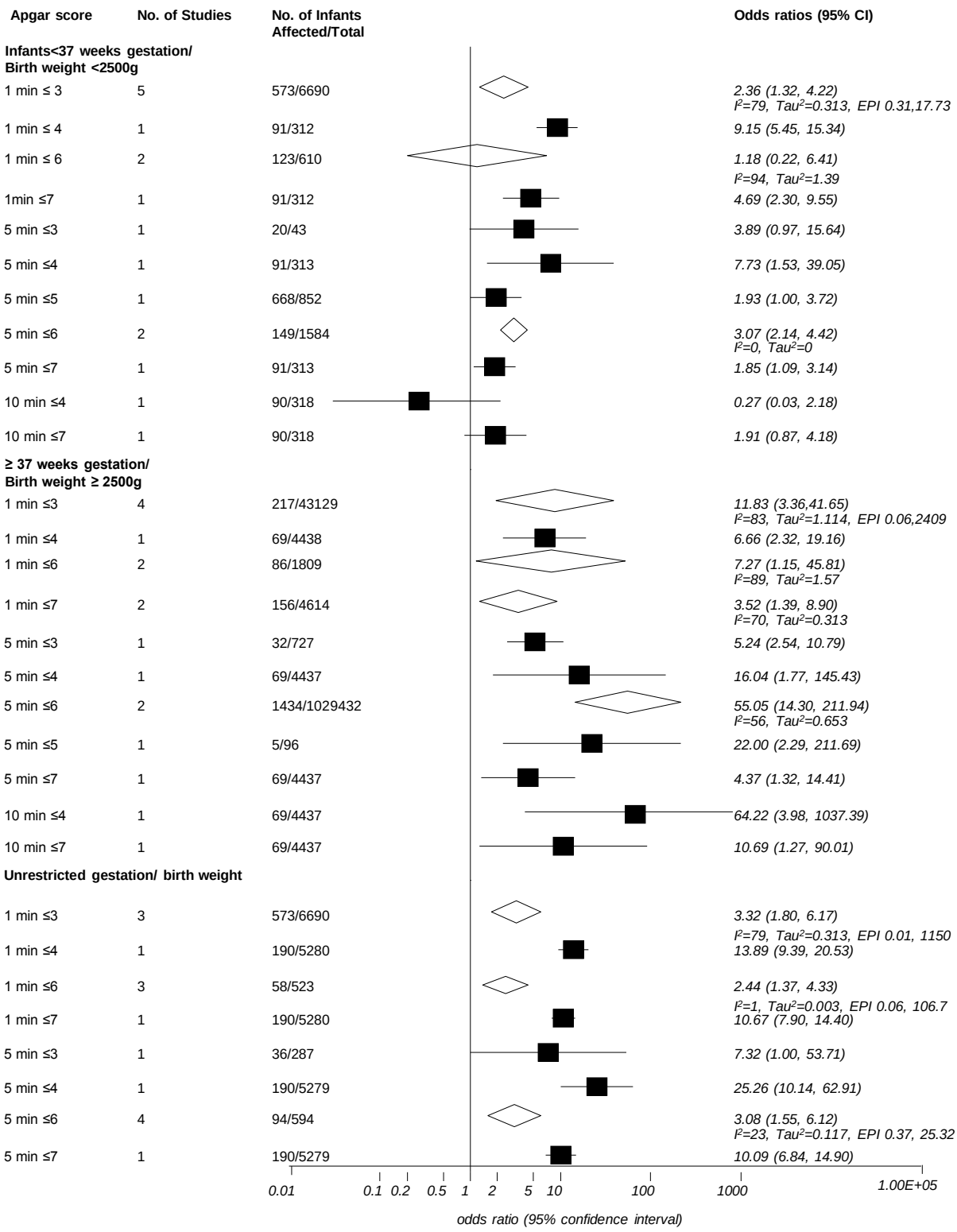


Figure 7.6 Forest plot of odds ratios for the association between Apgar score and infant mortality. Single studies are represented by a filled box, and pooled studies (and 95% confidence intervals) by a diamond

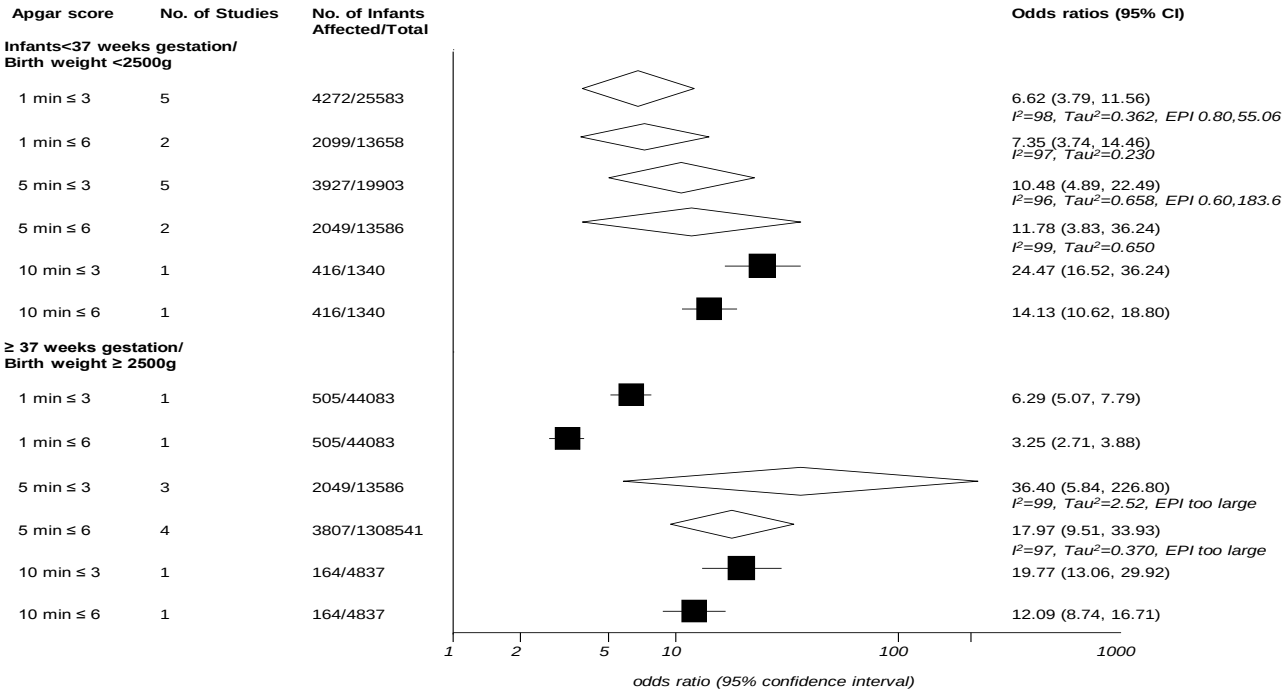


Figure 7.7 Forest plot of odds ratios for the association between Apgar score and cerebral palsy. Single studies are represented by a filled box, and pooled studies (and 95% confidence intervals) by a diamond

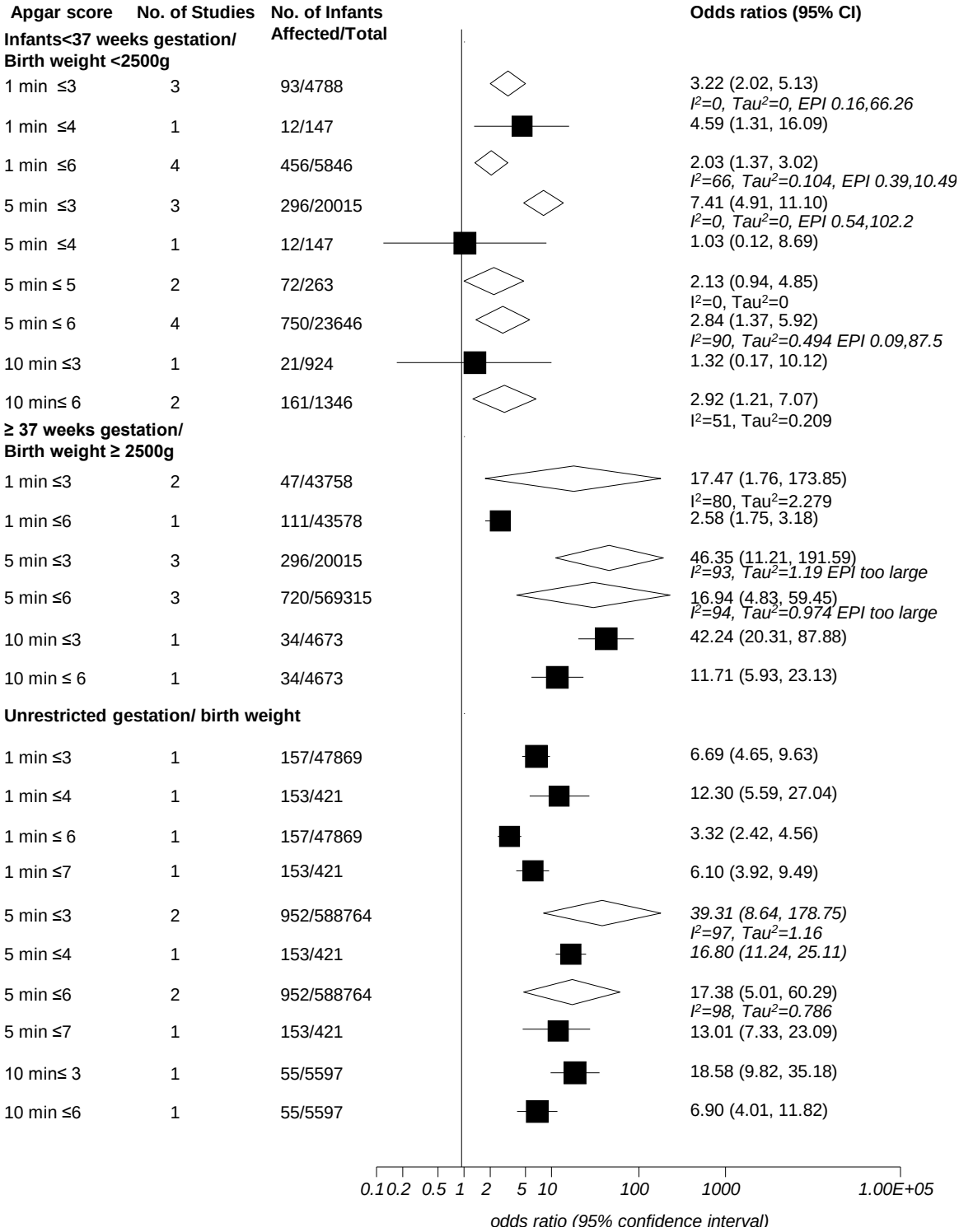


Figure 7.8 Forest plot of odds ratios for the association between Apgar score and childhood morbidity. Single studies are represented by a filled box, and pooled studies (and 95% confidence intervals) by a diamond

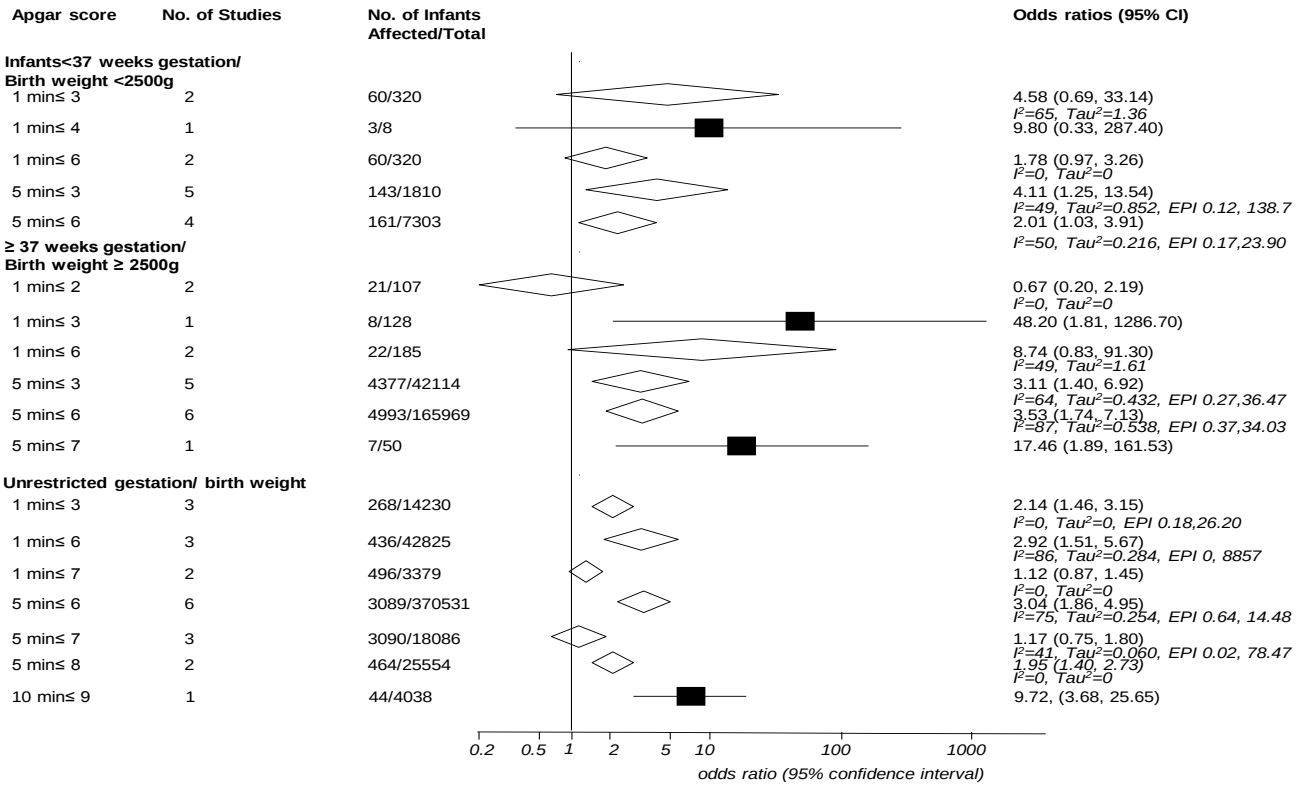


Table 7.1 Methodological quality of studies included in systematic review of Apgar score and adverse outcomes

Quality Item	Number (%) of studies n=87 Yes	No	Unclear
Cohort study design	70 (80)	16 (19)	1 (1)
Population adequately described	80 (92)	3 (3)	4 (5)
Consecutive recruitment	47 (54)	10 (12)	30 (34)
Prospective recruitment	28 (33)	53 (60)	6 (7)
Appropriate outcome measure	87 (100)	0	0
Outcome measure blinded	3 (3)	1 (1)	83 (96)
>90% of individuals had outcome measure	62 (72)	22 (25)	3 (3)
Index test and outcome measure described	8 (10)	71 (80)	8 (10)
Intervention between index test and outcome	2 (2)	0	85 (98)
Quality Classification			
High	48 (56)	-	-
Medium	34 (38)	-	-
Low	5 (6)	-	-

Table 7.2 Subgroup analysis according to Apgar score and outcome, where possible, for study quality, year of birth of study population, place of study and exclusion of congenital anomalies

Apgar score	Population	Number of studies	Subgroup	Odds ratio (95% CI)	Estimated prediction interval (EPI)	I^2 Tau^2
Neonatal death						
1 min ≤ 3	Preterm	4	High quality studies	6.23 (3.31, 11.74)	0.38,103.57	I^2=79, Tau^2=0.32
1 min ≤ 3	Preterm	4	Congenital anomalies excluded	11.79(4.63, 30.05)	0.14, 1010.75	I^2=96, Tau^2=0.84
1 min ≤ 3	Preterm	2	Year of birth ≥ 1990	7.75 (0.54, 111.16)	-	I^2=98, Tau^2=3.63
1 min ≤ 6	Preterm	2	High quality studies	7.72 (5.35, 11.14)	-	I^2=0, Tau^2=0
1 min ≤ 6	Preterm	3	Congenital anomalies excluded	16.01 (11.37, 22.55)	0.72, 357.11	I^2=31, Tau^2=0.03
5 min ≤ 3	Preterm	4	High quality studies	35.55 (8.11, 155.89)	0.03,48460	I^2=99, Tau^2=2.24
5 min ≤ 3	Preterm	4	Congenital anomalies excluded	33.52 (6.66, 168.79)	0.01, 90443	I^2=99, Tau^2=2.69
5 min ≤ 3	Preterm	3	Year of birth ≥ 1990	32.74 (4.19, 255.63)	Too large	I^2=99, Tau^2=3.28
5 min ≤ 6	Preterm	6	High quality studies	16.00 (4.51, 56.80)	0.14, 1787.33	I^2=99, Tau^2=2.47
5 min ≤ 6	Preterm	7	Congenital anomalies excluded	14.58 (4.95, 14.92)	0.28, 765.08	I^2=99, Tau^2=2.07
5 min ≤ 6	Preterm	4	Year of birth ≥ 1990	16.63 (3.81, 72.62)	0.01, 22252	I^2=99, Tau^2=2.23
5 min ≤ 6	Preterm	9	USA/Europe/ Australia/NZ studies	17.18 (6.41, 46.01)	0.41, 713.13	I^2=99, Tau^2=2.23
1 min ≤ 3	Term	2	High quality studies	104.35 (59.73, 102.31)	-	I^2=0, Tau^2=0
1 min ≤	Term	2	Congenital	44.39	-	I^2=99,

3			anomalies excluded	(1.10, 1791.77)		Tau^2=7.07
1 min ≤ 6	Term	2	High quality studies	61.86 (36.11, 105.96)	-	I^2=0, Tau^2=0
1 min ≤ 6	Term	2	Congenital anomalies excluded	21.36 (2.03, 225.2)	-	I^2=98, Tau^2=2.84
5 min ≤ 3	Term	2	High quality studies	179.42 (38.17, 843.5)	-	I^2=62, Tau^2=0.89
5 min ≤ 3	Term	3	Congenital anomalies excluded	247.30 (36.18, 1644.78)	Too large	I^2=98, Tau^2=2.76
5 min ≤ 6	Term	2	High quality studies	87.86 (57.23, 134.87)	-	I^2=24, Tau^2=0.04
5 min ≤ 6	Term	3	Congenital anomalies excluded	83.44 (22.65, 307.3)	Too large	I^2=97, Tau^2=1.29
5 min ≤ 6	Term	5	USA/Europe/ Australia/NZ studies	88.96 (34.58, 228.85)	2.7,3332.7	I^2=95, Tau^2=1.06
1 min ≤ 3	Unrestricted	3	High quality studies	16.91 (9.52, 30.06)	0.41, 703.14	I^2=0, Tau^2=0
1 min ≤ 3	Unrestricted	2	Year of birth ≥ 1990	59.36 (6.04, 583.58)	-	I^2=89, Tau^2=2.45
1 min ≤ 3	Unrestricted	7	USA/Europe/ Australia/NZ studies	27.91 (11.32, 68.85)	1.05, 739.44	I^2=98, Tau^2=1.41
1 min ≤ 6	Unrestricted	4	High quality studies	19.71 (10.66, 36.44)	1.74, 222.6	I^2=63, Tau^2=0.22
1 min ≤ 6	Unrestricted	3	Year of birth ≥ 1990	46.33 (19.22, 111.64)	Too large	I^2=96, Tau^2=0.45
1 min ≤ 6	Unrestricted	6	USA/Europe/ Australia/NZ studies	22.99 (10.35, 51.03)	1.27, 416.06	I^2=97, Tau^2=0.92
5 min ≤ 6	Unrestricted	5	High quality studies	109.68 (77.16, 155.88)	34.26, 351.05	I^2=90, Tau^2=0.10
5 min ≤ 6	Unrestricted	3	Year of birth ≥ 1990	115.25 (66.20, 204.17)	0.19, 70822	I^2=90, Tau^2=0.17

5 min ≤ 6	Unrestricted	5	USA/Europe/ Australia/NZ studies	97.74 (58.30, 163.85)	13.99, 682.84	I^2=95, Tau^2=0.30
Neonatal morbidity						
1 min ≤ 3	Preterm	3	High quality studies	3.50 (2.87, 4.28)	0.96, 12.77	I^2=0, Tau^2=0
1 min ≤ 3	Preterm	2	Congenital anomalies excluded	2.45 (1.52, 3.87)	-	I^2=2, Tau^2=0.001
1 min ≤ 3	Preterm	2	Year of birth ≥ 1990	3.53 (2.87, 4.34)	-	I^2=0, Tau^2=0
1 min ≤ 3	Preterm	2	USA/Europe/ Australia/NZ studies	2.13 (1.12, 4.05	0.12, 38.53	I^2=84, Tau^2=0.35
Infant mortality						
1 min ≤ 3	Preterm	3	Year of birth ≥ 1990	4.28 (2.78, 6.58)	0.02, 1038.43	I^2=96, Tau^2=0.14
1 min ≤ 3	Preterm	3	Congenital anomalies excluded	9.34 (3.14, 27.73)	Too large	I^2=98, Tau^2=0.84
5 min ≤ 3	Preterm	2	Congenital anomalies excluded	12.97 (1.39, 121.3)	-	I^2=95, Tau^2=2.47
5 min ≤ 3	Preterm	4	Year of birth ≥ 1990	7.90 (4.24, 14.73)	0.52,119.55	I^2=90, Tau^2=0.30
5 min ≤ 3	Preterm	4	USA/Europe/ Australia/NZ studies	12.68 (5.52, 29.10)	0.25, 639.57	I^2=97, Tau^2=0.65
Childhood morbidity						
5 min ≤ 3	Preterm	3	High quality studies	5.98 (1.61, 22.27)	Too large	I^2=48, Tau^2=0.67
5 min ≤ 3	Preterm	3	Congenital anomalies excluded	3.98 (1.80, 8.80)	0.02, 683.7	I^2=0, Tau^2=0
5 min ≤ 3	Preterm	3	USA/Europe/ Australia/NZ studies	2.64 (0.66, 10.63)	Too large	I^2=41, Tau^2=0.66
5 min ≤ 3	Term	4	Congenital anomalies excluded	4.14 (1.97, 8.69)	0.49,35.07	I^2=17, Tau^2=0.10
5 min ≤ 3	Term	4	USA/Europe/ Australia/NZ studies	2.44 (1.22, 4.85)	0.17, 34.02	I^2=57, Tau^2=0.25
5 min ≤	Term	2	High quality	4.36	-	I^2=29,

6			studies	(2.89, 6.56)		Tau^2=0.03
5 min ≤ 6	Term	5	Congenital anomalies excluded	4.38 (3.19, 6.01)	2.62, 7.33	I^2=0, Tau^2=0
5 min ≤ 6	Term	5	USA/Europe/ Australia/NZ studies	3.29 (1.59, 6.78)	0.24, 44.49	I^2=89, Tau^2=0.53
5 min ≤ 6	Unrestricted	3	High quality studies	3.05 (1.35, 6.90)	0, 35355	I^2=78, Tau^2=0.37
5 min ≤ 6	Unrestricted	2	Congenital anomalies excluded	3.20 (0.95, 10.80)	-	I^2=86, Tau^2=0.67
5 min ≤ 6	Unrestricted	2	Year of birth ≥ 1990	2.20 (1.22, 3.95)	-	I^2=0, Tau^2=0
5 min ≤ 6	Unrestricted	4	USA/Europe/ Australia/NZ studies	3.70 (2.03, 6.74)	0.28,48.88	I^2=78, Tau^2=0.27

Table 7.3 The predictive ability of the Apgar score for neonatal mortality

Apgar score	No. of studies	Sensitivity (95% CI)	Specificity (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
Unrestricted population					
1 min =0	1	0.07 (0.05-0.1)	0.99 (0.998-0.999)	49.88 (31.52-78.95)	0.93 (0.90-0.95)
1 min ≤ 1	1	0.36 (0.32-0.41)	0.98 (0.982-0.985)	21.93 (18.77-25.63)	0.65 (0.60-0.70)
1 min ≤ 2	1	0.52 (0.47 to 0.57)	0.96 (0.962 to 0.966)	14.43 (12.93-16.10)	0.50 (0.45 to 0.55)
1 min ≤ 3	8	0.45 (0.36 to 0.53)	0.97 (0.95 to 0.98)	16.17 (15.22 to 52.87)	0.57 (0.49 to 0.66)
1 min ≤ 4	1	0.67 (0.61 to 0.70)	0.90 (0.90 to 0.91)	6.85 (6.34 to 7.41)	0.38 (0.33 to 0.43)
1 min ≤ 5	1	0.72 (0.68 to 0.77)	0.86 (0.86 to 0.87)	5.31 (4.97 to 5.67)	0.32 (0.27 to 0.37)
1 min ≤ 6	8	0.71 (0.67 to 0.75)	0.90 (0.87 to 0.93)	7.75 (5.77 to 10.42)	0.31 (0.28 to 0.36)

1 min ≤ 7	2	0.79 (0.68 to 0.86)	0.77 (0.66 to 0.85)	3.34 (2.21 to 5.23)*	0.28 (0.19 to 0.43)*
1 min ≤ 8	1	0.91 (0.88 to 0.94)	0.46 (0.46 to 0.47)	1.69 (1.64 to 1.75)	0.19 (0.14 to 0.26)
1 min ≤ 9	1	0.98 (0.97 to 0.99)	0.14 (0.13 to 0.14)	1.14 (1.12 to 1.15)	0.12 (0.06 to 0.15)
5 min ≤ 3	4	0.29 (0.16 to 0.46)	0.999 (0.997 to0.999)	212.9 (141.02 to 321.27)	0.71 (0.58 to 0.88)
5 min ≤ 6	7	0.53 (0.42 to 0.63)	0.99 (0.98 to 0.99)	48.43 (28.91 to 81.13)	0.48 (0.38 to 0.59)
5 min ≤ 7	1	0.07 (0.05 to 0.10)	0.994 (0.992 to 0.995)	11.50 (7.65 to 17.28)	0.93 (0.91 to 0.95)
10 min ≤ 3	1	0.28 (0.23 to 0.34)	0.999 (0.999 to 0.999)	444.13 (355.74 to 554.48)	0.72 (0.67 to 0.78)
10 min ≤ 6	1	0.42 (0.36 to 0.49)	0.998 (0.998 to 0.998)	226.17 (193.87 to 263.84)	0.58 (0.52 to 0.64)
Term/ normal birth weight population					
1 min =0	1	0.11 (0.06 to 0.18)	0.999 (0.998 to 0.999)	74.28 (39.69 to 139.0)	0.90 (0.84 to 0.95)
1 min ≤ 1	2	0.37 (0.30 to 0.46)	0.99 (0.98 to 0.99)	44.86 (13.70 to 146.94)*	0.63 (0.56 to 0.72)*
1 min ≤ 2	2	0.47 (0.39 to 0.55)	0.98 (0.96 to 0.99)	24.13 (8.43 to 69.11)*	0.55 (0.47 to 0.64)*
1 min ≤ 3	7	0.47 (0.38 to 0.56)	0.98 (0.96 to 0.99)	23.33 (10.75 to 50.65)	0.54 (0.46 to 0.64)
1 min ≤ 4	2	0.59 (0.50 to 0.67)	0.95 (0.89 to 0.98)	11.54 (3.46 to 38.48)*	0.45 (0.37 to 0.55)*
1 min ≤ 5	2	0.63 (0.55 to 0.71)	0.92 (0.84 to 0.97)	8.81 (2.51 to 30.93)	0.42 (0.34 to 0.52)
1 min ≤ 6	6	0.66 (0.61 to 0.70)	0.95 (0.91 to 0.97)	12.73 (7.26 to 22.31)	0.36 (0.31 to 0.42)
1 min ≤ 7	3	0.78 (0.71 to 0.83)	0.80 (0.72 to 0.86)	3.82 (2.41 to 6.05)*	0.31 (0.22 to 0.44)*
1 min ≤ 8	2	0.95 (0.45 to 1.00)	0.44 (0.41 to 0.48)	1.65 (1.58 to 1.73)*	0.20 (0.04 to 0.95)*
5 min ≤ 1	1	0.30 (0.13 to 0.53)	1.00 (0.99 to 1.00)	1537.1 (337.1 to 7008.9)	0.70 (0.53 to 0.91)
5 min ≤ 2	1	0.30 (0.13 to 0.53)	0.999 (0.998 to 1.00)	384.28 (151.89 to 972.21)	0.70 (0.53 to 0.91)

5 min ≤ 3	5	0.26 (0.16 to 0.38)	0.999 (0.997 to 0.999)	188.41 (75.73 to 468.77)	0.74 (0.64 to 0.86)
5 min ≤ 4	1	0.44 (0.23 to 0.66)	0.998 (0.996 to 0.998)	182.99 (99.05 to 388.08)	0.57 (0.40 to 0.81)
5 min ≤ 5	1	0.57 (0.35 to 0.77)	0.996 (0.994 to 0.997)	126.87 (79.93 to 201.38)	0.44 (0.27 to 0.70)
5 min ≤ 6	6	0.50 (0.36 to 0.64)	0.98 (0.95 to 1.00)	30.97 (10.86 to 88.36)	0.51 (0.38 to 0.67)
5 min ≤ 7	1	0.78 (0.56 to 0.93)	0.98 (0.97 to 0.98)	30.76 (24.03 to 39.37)	0.22 (0.10 to 0.48)
5 min ≤ 8	1	0.87 (0.66 to 0.97)	0.89 (0.88 to 0.90)	7.90 (6.68 to 9.34)	0.15 (0.05 to 0.42)
10 min ≤ 3	1	0.32 (0.24 to 0.42)	1.00 (1.00 to 1.00)	958.99 (698.89 to 1315.9)	0.68 (0.59 to 0.77)
10 min ≤ 6	1	0.47 (0.37 to 0.47)	0.999 (0.999 to 0.999)	411.57 (329.95 to 513.39)	0.53 (0.45 to 0.64)
Preterm/ low birth weight population					
1 min =0	1	0.06 (0.04 to 0.09)	0.998 (0.995 to 0.999)	32.24 (11.04 to 94.15)	0.94 (0.91 to 0.97)
1 min ≤ 1	1	0.36 (0.31 to 0.42)	0.96 (0.95 to 0.97)	9.16 (7.08 to 11.85)	0.67 (0.61 to 0.72)
1 min ≤ 2	1	0.54 (0.48 to 0.59)	0.93 (0.91 to 0.94)	7.22 (6.02 to 8.67)	0.50 (0.44 to 0.56)
1 min ≤ 3	9	0.54 (0.43 to 0.64)	0.88 (0.82 to 0.92)	4.47 (3.02 to 6.62)	0.53 (0.43 to 0.64)
1 min ≤ 4	2	0.69 (0.64 to 0.74)	0.77 (0.62 to 0.88)	3.62 (2.09 to 6.25)*	0.37 (0.31 to 0.43)*
1 min ≤ 5	1	0.76 (0.71 to 0.81)	0.79 (0.77 to 0.81)	3.66 (3.30 to 4.07)	0.30 (0.25 to 0.37)
1 min ≤ 6	6	0.82 (0.74 to 0.88)	0.71 (0.62 to 0.79)	2.87 (2.23 to 3.69)	0.25 (0.18 to 0.34)
1 min ≤ 7	2	0.81 (0.70 to 0.88)	0.66 (0.56 to 0.75)	2.34 (1.84 to 2.98)*	0.30 (0.19 to 0.46)*
1 min ≤ 8	1	0.93 (0.90 to 0.96)	0.35 (0.33 to 0.37)	1.44 (1.38 to 1.51)	0.19 (0.13 to 0.29)
1 min ≤ 9	1	0.99 (0.97 to 1.00)	0.10 (0.09 to 0.11)	1.09 (1.07 to 1.11)	0.13 (0.05 to 0.35)
5 min ≤ 3	7	0.41 (0.32 to 0.52)	0.98 (0.94 to 0.99)	18.19 (8.29 to 39.91)	0.60 (0.51 to 0.71)
5 min ≤ 4	1	0.67 (0.22 to 0.96)	0.92 (0.88 to 0.96)	8.71 (4.21 to 18.02)	0.36 (0.12 to 1.12)

5 min ≤ 5	1	0.31 (0.23 to 0.40)	0.94 (0.92 to 0.95)	4.90 (3.33 to 7.19)	0.74 (0.65 to 0.83)
5 min ≤ 6	10	0.63 (0.53 to 0.71)	0.90 (0.80 to 0.96)	6.58 (3.10 to 14.00)	0.41 (0.33 to 0.52)
10 min ≤ 3	1	0.25 (0.19 to 0.33)	0.995 (0.994 to 0.996)	50.01 (36.71 to 68.13)	0.75 (0.69 to 0.82)
10 min ≤ 6	1	0.40 (0.32 to 0.48)	0.987 (0.986 to 0.988)	30.49 (24.60 to 37.77)	0.61 (0.54 to 0.69)

*= Values calculated using Meta-Disc software

Table 7.4 The predictive ability of the Apgar score for cerebral palsy

Apgar score	No. of studies	Sensitivity (95% CI)	Specificity (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
Unrestricted population					
1 min ≤ 3	1	0.25 (0.18 to 0.32)	0.95 (0.95 to 0.96)	5.28 (4.01 to 6.95)	0.79 (0.72 to 0.86)
1 min ≤ 4	1	0.28 (0.21 to 0.35)	0.97 (0.94 to 0.99)	9.20 (4.43 to 19.07)	0.75 (0.68 to 0.83)
1 min ≤ 6	1	0.43 (0.35 to 0.51)	0.82 (0.81 to 0.82)	2.33 (1.94 to 2.80)	0.70 (0.61 to 0.80)
1 min ≤ 7	1	0.60 (0.51 to 0.67)	0.81 (0.75 to 0.85)	3.07 (2.32 to 4.04)	0.50 (0.41 to 0.62)
5 min ≤ 3	2	0.09 (0.06 to 0.15)	0.997 (0.986 to 0.999)	35.19 (7.40 to 167.32)*	0.90 (0.84 to 0.97)*
5 min ≤ 4	1	0.46 (0.37 to 0.56)	0.95 (0.94 to 0.96)	9.46 (7.39 to 12.12)	0.56 (0.47 to 0.67)
5 min ≤ 6	2	0.21 (0.16 to 0.27)	0.98 (0.95 to 0.99)	13.68 (3.77 to 49.61)	0.80 (0.76 to 0.85)
5 min ≤ 7	1	0.48 (0.40 to 0.57)	0.93 (0.90 to 0.96)	7.20 (4.48 to 11.58)	0.55 (0.47 to 0.65)
10 min ≤ 3	1	0.26 (0.15 to 0.39)	0.98 (0.98 to 0.99)	14.11 (8.62 to 23.08)	0.76 (0.65 to 0.89)
10 min ≤ 6	1	0.44 (0.30 to 0.58)	0.90 (0.89 to 0.91)	4.32 (3.17 to 5.89)	0.63 (0.50 to 0.79)
Pre-term/ low birth weight					
5 min ≤ 3	3	0.07 (0.03 to 0.13)	0.99 (0.98 to 0.99)	6.92 (4.74 to 10.10)	0.95 (0.93 to 0.97)
Term/ normal birth weight					
1 min ≤ 3	2	0.28 (0.22 to 0.35)	0.96 (0.95 to 0.96)	13.22 (1.13 to 154.22)*	0.74 (0.65 to 0.85)*
5 min ≤ 3	3	0.10 (0.07	0.998	40.95 (9.53	0.90 (0.86

		to 0.14)	(0.991 to 0.999)	to 175.92)*	to 0.94)*
5 min ≤ 6	3	0.20 (0.17 to 0.23)*	0.992 (0.992 to 0.992)*	8.65 (2.04 to 36.80)*	0.82 (0.79 to 0.85)*
10 min ≤ 3	1	0.38 (0.22 to 0.56)	0.99 (0.98 to 0.99)	26.47 (16.24 to 43.17)	0.63 (0.48 to 0.82)
10 min ≤ 6	1	0.50 (0.32 to 0.68)	0.92 (0.91 to 0.93)	6.36 (4.48 to 9.02)	0.54 (0.39 to 0.76)

*= Values calculated using Meta-Disc software

7.5 Discussion

A low Apgar score showed a strong, consistent association with neonatal mortality. The relationship was highest at lower thresholds, and generally decreased (but remained strong) as the threshold increased. The fact that the association remained significant at a 1 minute Apgar thresholds of 8 and 9 was surprising, but is likely to reflect that these cut offs include all of the lower scores which are very strongly associated with poor outcome, as the higher scores were not considered in isolation. A lower score at later time intervals was most strongly associated with this outcome, which reflects the fact that babies who are in poorer condition for longer are more likely to have undergone a significant pathological process causing compromise, increasing the risk of death. The relationship between a low Apgar score and neonatal mortality appeared stronger in a term (≥ 37 weeks gestation) or normal birth weight (≥ 2.5kg) population than a pre-term or unrestricted population. For a term population, the association between a low Apgar score and neonatal morbidity, infant mortality, childhood cerebral palsy and other childhood morbidity, was significant at all but one threshold of the Apgar score assessed. The magnitude of the association was less than that for neonatal mortality. However, the

association between a low Apgar score and other adverse outcomes in a pre-term or low birth weight population was less consistent. This is likely to reflect the fact that a low Apgar score in pre-term infants may reflect the differing physiology of infants born prior to term, and be low as a result of immaturity rather than a pathological process with sequelae. Also, this population is at higher risk of adverse outcomes related to prematurity, and as such the condition at birth may not reflect their overall risk of developing complications, thus diluting the relationship.

Strengths and limitations of this review

The strengths of this review lie in the methodology used. Extensive literature searches were performed without language restrictions, and every effort was made to search for unpublished data. The review complies with the most recent guidance on performing and reporting of systematic reviews of observational studies.[78;84] The most up to date techniques have been used for performing and interpreting meta-analysis.[102;130;131]

However, despite attempts to include all available data, Peter's test suggested the presence of a small study effect in the only meta-analysis group large enough for this to be performed. Another limitation is the varying number of studies that contributed to each analysis. The majority of studies used an Apgar score threshold of ≤ 3 or ≤6 to define a low score, with only 2 papers providing data on the Apgar score at each threshold (Apgar and James 1962 and Jennett et al 1981), therefore direct comparisons were only possible for these; others are indirect and the magnitude in association may be affected by differences in population or other study characteristics.

The Apgar score at 1, 5 and 10 minutes in a single infant are not independent, i.e. an infant with a low score at 10 minutes is likely to have also had a low score at 1 and 5 minutes. Due to the nature of the reporting in the primary studies, where multiple time points were reported, it was not possible to analyse the outcome following a low score at more than one time point e.g. less than 5 at both 1 and 5 minutes. However, by only performing meta-analysis for individual thresholds at a particular time point (as reported in the primary studies), and only including each study once in any given meta-analysis, care was taken to ensure that individuals were not counted multiple times in any analysis.

The reporting quality of included studies was particularly poor regarding the execution of the index test, i.e. very few provided descriptions of how the score was assigned or the personnel performing this. Given the subjective nature of some components of the score, the inter-observer variability is an important consideration of this test, but none of the included studies reported this. O'Donnell et al assessed the inter-observer variability of the 5 minute Apgar score through showing video clips to observers and comparing these with scores assigned by the member of staff attending the delivery.[141] They found that scores assigned varied widely, both between the observers watching the video clip and between the scores actually assigned at delivery and the video scores, with a mean difference of 2.4.[141] In another study by Clark and Hakansen, inter-rater reliability was found to be 68% among paediatric staff, but 24% among community hospital nurses.[142]

It was not possible to investigate within this review whether different components of the Apgar score were particularly correlated with adverse outcome. Only two of the included studies reported on components of the Apgar score, one found that a greater proportion of infants with a 1 minute Apgar score ≤6 who died had a heart rate below 100 (72.5% of infants that died versus 33% of survivors)(Colburn 1960). The other study found that heart rate and reflex irritability at 1 minute were the best discriminators between babies that were 'healthy or relatively healthy' and 'severely ill' in the neonatal period (Valentin et al 1993).

Heygi et al investigated the relationship between components of the 1 minute Apgar score in pre-term infants, and found that the strongest correlations were between respiratory rate, tone and reflex.[143] The heart rate corresponded less well with other components, and colour had the lowest correlation. They constructed a model to compare the contribution of each factor to the 1 minute Apgar score from 0-10, and found that respiratory rate, tone and reflex produced the largest increase in total score, followed by heart rate and colour, reflecting the large contribution of heart rate at lower Apgar scores and smaller contribution at higher scores. Tone contributed more to higher scores than heart rate in this population.[143]

Subgroup analyses were planned within this review to assess the implications of a low Apgar score in pre-term and term born populations separately, however due to the fact that many studies reported birth weight rather than gestational age in the population characteristics, a broader subgroup of infants born at less than 37 weeks gestation, or of birth weight less than 2.5 kg, was used. This may

therefore lead to some crossover between pre-term and small for gestational age infants in both subgroups and affect the extrapolation of the results to clinical practice. Other potential factors which may affect either the Apgar score of an individual, or the relationship between a low score and adverse outcome, were addressed. Subgroup analyses for study quality, location, year of birth ≥ 1990 and exclusion of congenital anomalies were performed. In general, the subgroup analyses performed did not affect the significance or direction of the results. In the cases where the results did change, this most commonly occurred in the pre-term population, where limiting the analysis according to these factors made the association between low Apgar score and adverse outcome non-significant. However, the results were inconsistent and the same subgroup analysis did not produce the same effect at all thresholds of the Apgar score examined. Due to the reporting quality of the primary studies, it was not possible to address the effect of other factors which may affect the Apgar score, such as medication administered to the mother.

Comparison with other studies

The only other systematic review reporting the association between a low Apgar score and neonatal mortality and cerebral palsy found a significant association between a low score and these outcomes.[77] The current review supports these findings, and has performed a more in depth investigation into these relationships at different thresholds and subgroups.

Implications for clinical practice

Meta-analysis confirms that a low Apgar score has a strong prognostic association with neonatal mortality, particularly in a population born at term. The association between a low Apgar score and neonatal morbidity, childhood cerebral palsy and other morbidities was also significant, suggesting that the Apgar score is a valid measure of neonatal wellbeing that has implications for long term health. However, no single threshold showed both high sensitivity and specificity for the adverse outcomes examined, therefore it is not possible to choose a standard cut-off that should be employed in clinical practice.

Recommendations

Future research is necessary to further explore subgroups to allow prognostication of a low Apgar score at an individual level to facilitate counselling of parents and target observation and intervention in the neonatal period. This could be performed through an individual patient data meta-analysis.[121] Another option would be to perform a large prospective cohort study in which further risk factors could be recorded, or individual components of the Apgar score used to predict outcome.

Finally, it is likely that more accurate predictions could be made using the Apgar score in combination with other tests of neonatal wellbeing that are performed at birth.

7.6 Conclusion

A low Apgar score at birth is strongly associated with neonatal mortality, morbidity and childhood cerebral palsy, particularly in a population born at term,

or with normal birth weight. Further research is required to identify the subgroups in which Apgar score may best predict adverse outcome. This may be done through individual patient data meta-analysis. Prognostic models also containing other factors may optimise the prediction of adverse outcome on an individual level. Adherence of future studies to the STARD criteria will facilitate meta-analysis.

PART B: SUMMARY OF EXISTING LITERATURE FOR NEONATAL HYPOTHERMIA TO TREAT HYPOXIA, AND DECISION-ANALYTIC MODELLING

CHAPTER 8: SUMMARY OF THE EXISTING LITERATURE TO SUPPORT THERAPEUTIC NEONATAL HYPOTHERMIA TO PREVENT SEQUELAE FOLLOWING HYPOXIA

8.1 Introduction

Peripartum asphyxia affects 3 to 5 per 1000 live births in developed countries, with moderate or severe hypoxic ischaemic encephalopathy occurring in 0.5 to 1 per 1000.[12] Moderate encephalopathy carries a 10 percent risk of death, and 30 percent risk of disability, and where encephalopathy is severe, 60 percent die and most survivors have neurological abnormalities.[144] The Sarnat grading of severity of encephalopathy is given in Table 8.1.[144] At a tissue level, hypoxic ischaemia initiates energy depletion, the accumulation of extracellular glutamate, and activation of receptors, loss of membrane homeostasis and increases in intracellular calcium and osmotic dysregulation, resulting in cell death.[11;144] It is thought that the process of cellular damage is an evolving course, and that although neurons may die during the initial event, many recover and undergo a latent phase, only to die hours or days later (secondary or delayed cell death).[145] The mechanisms of delayed neuronal death include hyperaemia, cytotoxic oedema, mitochondrial failure, accumulation of excitotoxins, nitric oxide synthesis, free radical damage and cytotoxic actions of activated microglia.[146]

Initial studies of intentionally inducing hypothermia in neonates were performed in the 1940s and 50s. The purpose of the treatment at this time was to resuscitate asphyxiated infants. Animal studies and a small case series of severely asphyxiated human infants showed successful resuscitation and increased survival in subjects who had undergone whole body cooling.[147] However, controversy existed due to data published on the adverse effects of neonatal exposure to cold,[148;149] and alternative techniques for resuscitation developed, therefore it was not until recent years that neonatal therapeutic hypothermia to prevent neurologic injury has gained credence and become an established treatment.

Table 8.1 Sarnat staging of encephalopathy

Sarnat Stage 1 (Mild)	**Sarnat Stage 2 (Moderate)**	**Sarnat Stage 3 (Severe)**
Hyperalert	Lethargic	Stuporous
Normal tone	Mild hypotonia	Flaccid
Overactive stretch reflexes	Overactive stretch reflexes	Decreased or absent stretch reflexes
Weak suck	Weak/absent suck	Absent suck
No seizures	Common, focal or multifocal	Uncommon
Less than 24 hours duration	2-14 days	Hours to weeks

8.2 Summary of the evidence to support neonatal hypothermia

8.2.1 Cochrane review: 'Cooling for newborns with hypoxic ischaemic encephalopathy'

A Cochrane review published in 2008 included randomised controlled trials comparing the use of therapeutic hypothermia with standard care in encephalopathic newborns, with evidence of peripartum asphyxia and without major congenital anomalies. The primary outcome measure was death or long term major neurodevelopmental disability.[150] Eight studies were included, comprising 638 infants born at near term gestation. Evidence of perinatal asphyxia to satisfy the inclusion criteria for the review was one of: Apgar score ≤ 5 at 10 minutes, mechanical ventilation or resuscitation at 10 minutes, or cord pH <7.1 or arterial pH <7.1 or base deficit of ≥ 12 within 60 minutes of birth. Evidence of encephalopathy was according to Sarnat staging. Method of therapeutic hypothermia was either by whole body or selective head cooling. Four trials performed selective head cooling with mild systemic hypothermia, meta-analysis showed that the risk ratio (RR) for mortality was 0.83 (95% CI 0.59 to 1.16),[151-154] compared to four trials that performed whole body cooling (RR 0.66, 95% CI 0.47 to 0.93).[155-158] For the outcome of major neurodevelopmental disability, three studies were included in the meta-analysis. The relative risk for this outcome was 0.73 (95% CI 0.53 to 0.99).[151;153;156] Of these, one study used a pH threshold of ≤ 7.09,[151] and the others used <7.00 as one of the inclusion criteria. The overall conclusion of the review was that cooling reduces mortality without increasing major disability in survivors, with the benefits outweighing short-term adverse effects. However, it was known that further trials were ongoing when the review was published, and it was felt that clarification regarding the best

method of administering hypothermia (selective head cooling versus mild systemic hypothermia), were required before recommending the treatment for clinical practice.

8.2.2 The TOBY trial (Whole body hypothermia for the treatment of perinatal asphyxial encephalopathy)

The TOBY trial, published in 2009, randomised infants who were at least 36 weeks gestation, less than six hours of age, with evidence of asphyxia defined as an Apgar score of 5 or less at 10 minutes of age, or umbilical cord, arterial or capillary pH of <7.00 or base deficit of ≤ 16mmol/L within 60 minutes of birth. Included infants had to have moderate- to-severe encephalopathy and abnormal background activity of at least 30 minutes duration or seizures on amplitude integrated electroencephalography (aEEG). Infants with congenital anomalies requiring surgery, or chromosomal anomalies involving brain dysgenesis, were excluded.[159] This was the largest randomised controlled trial to date of hypothermia for perinatal encephalopathy. Infants assigned to the treatment arm had hypothermia maintained by a cooling blanket, with a target rectal temperature of 33-34°C. The period of cooling was for 72 hours after randomisation. Temperature was then returned to normal in a controlled fashion, and neonates were monitored with daily cranial ultrasound scans for the first four days, and magnetic resonance imaging (MRI) within 5-14 days of birth. Infants were followed up until 18 months of age, at which point they underwent a structured neurologic examination, and the presence and type of cerebral palsy determined. The Bayley scales of Infant Development II (BSID-II) were also applied, and adverse outcomes prior to this point recorded. The main findings were that infants who were cooled and survived had a reduced risk of

cerebral palsy (RR 0.67, 95% CI 0.47 to 0.96), and improved scores of the Mental and Psychomotor development Index of the BSID-II.

Following the publication of this trial, neonatal hypothermia for the treatment of hypoxic-ischaemic encephalopathy has become widespread in UK practice; the criteria most commonly used to determine treatment, and the type and duration of cooling are those used in the TOBY protocol as described above.[58]

8.2.3 Meta-analysis of Randomised Controlled Trials 2010

Following publication of the TOBY trial results, a meta-analysis[160] was published comprising three trials, with data on neurological outcomes up to at least 18 months of age, and similar entry criteria.[153;156;159] A total of 767 infants were included of which 518 survived: fixed effects meta-analysis demonstrated a reduced risk of cerebral palsy in survivors (RR 0.69, 95% CI 0.54 to 0.89). The authors felt that the results strongly supported the use of therapeutic hypothermia in neonates with hypoxic ischaemic encephalopathy, and that any remaining ongoing studies would have to show very large adverse effects to change the results, given that the current studies are homogeneous in their findings favouring treatment.

8.2.4 Neo.nEURO.Network RCT 2010

A further randomised controlled trial, conducted in central Europe, allocated infants to systemic hypothermia based on the same entry criteria as the TOBY trial. In addition to systemic hypothermia for 72 hours, they also administered 0.1mg/kg of morphine to all participants every four hours, with the rationale of reducing discomfort attributable to encephalopathy and counteracting the stress response induced by hypothermia.[161] 129 neonates were included, recruitment was stopped early due to

ethical concerns about withholding treatment in the control group. The results favoured treatment when the composite adverse outcome of death or severe disability (OR 0.21, 95% CI 0.09 to 0.54) was considered, and for cerebral palsy in survivors (OR 0.15, 95% CI 0.04 to 0.60). The apparent greater magnitude of effect seen in this study was thought by the authors to be potentially due to the smaller sample size in this study, or the possible added benefit of administering opioids in addition to hypothermia.

8.2.5 ICE (Infant Cooling Evaluation) trial

This randomised controlled trial was conducted in neonatal intensive care units in Australia, New Zealand, Canada and the United States and published in 2011.[162] Neonates of 35 weeks gestation or more with moderate or severe encephalopathy defined according to modified Sarnat criteria, and evidence of peripartum hypoxia-ischaemia (at least two of: Apgar score of 5 or less at 10 minutes, continued need for mechanical ventilation at 10 minutes, and/ or metabolic acidosis defined as pH <7.00 or base deficit of ≥ 12mmol/L within 60 minutes of birth) were included. Cooling was administered via refrigerated gel packs applied across the chest and/ or under the head and shoulders, to maintain the core temperature at 33-34°C. 221 infants were randomised. There was a statistically significant difference in the primary outcome of death or major disability (RR 0.77, 95% CI 0.62 to 0.98). There was no significant difference in rates of cerebral palsy at 2 years of age (RR 0.92, 95% CI 0.54 to 1.59).

8.2.6 Cost-effectiveness of therapeutic hypothermia to treat neonatal encephalopathy

Regier et al[163] published a cost-effectiveness analysis based on the effectiveness data of the three trials included in the meta-analysis summarised in section 8.2.3.[160] The cost data was obtained from the TOBY trial, as this was the only source of prospectively collected resource utilisation data for encephalopathic infants.[159] A decision-analytic model based analysis from the perspective of the NHS and personal social services was performed, and incremental effectiveness was estimated in terms of disability free life years (DFLY) gained. The time horizon was the first 18 months after birth.

The incremental cost per DFLY gained was £19,931. The baseline cost effectiveness acceptability curve showed that if the willingness to pay threshold is £30,000, there is a 69% probability that cooling is cost-effective, although this increased when the number of infants was increased to reflect the national incidence of encephalopathy, as per the UK Cooling Register, or the time horizon was lengthened.

8.3 Mechanisms of neuroprotection

A number of mechanisms by which hypothermia may improve outcome after hypoxic ischaemic insult have been suggested, and summarised by Drury et al.[145] The cerebral metabolism is reduced by hypothermia, delaying the onset of anoxic cell depolarisation, and cooling during the reperfusion phase may reduce free radicals and lipid peroxidation. This supports the use of passive cooling during the period of assessment before a decision for active cooling is made, but does not account for the protective effects of delayed, post insult cooling. A period of secondary

hypoperfusion in the latent phase following a cerebral insult has been noted. This was initially thought to be an adverse outcome, however evidence suggests that this effect is mediated by suppressed cerebral metabolism, and is associated with increased tissue oxygen levels. Prolonging the phase of secondary hypoperfusion through hypothermia has been shown in animal studies to improve neural outcome.

Hypothermia may additionally have a particular role in suppressing post-ischaemic cell death, through suppression of the Caspase pathway, a crucial component in apoptosis. Sheep studies have demonstrated suppression of activated Caspase-3 in association with hypothermic survival. Additionally, cooling potentially suppresses multiple aspects of the inflammatory cascade, which in non-cooled cases increases release of cytokines and interleukins, which may exacerbate injury through direct neurotoxicity and induction of apoptosis. It is likely to be the intracytoplasmic effects that are critical to protection.[145]

8.4 Potential side-effects of therapeutic hypothermia

Side effects include arrhythmias, most commonly sinus bradycardia. This represents a normal physiological response to hypothermia, however infants need to be monitored closely for hypoperfusion: the hypothermia in combination with cardiovascular effects resulting from the hypoxic insult may lead to the need for volume resuscitation and inotropic support.[144] Minor respiratory and cardiovascular effects were frequently observed in the TOBY trial.[159] Within the Cochrane review, there was a significantly higher risk of sinus bradycardia in the treatment groups of included studies (RR 5.96, 95% CI 2.15 to 16.49).[150] Coagulopathy is another potential side effect of hypothermia; this should be monitored during treatment and

blood products administered as necessary.[144] There was no significant difference between the treatment and control groups for this outcome in the Cochrane meta-analysis.[150] Fat necrosis is a rare side effect but infants should be closely monitored for skin changes suggestive of this, and their position changed regularly while cooling is underway.[144] No significant differences in severe adverse outcomes between the cooling and the standard care arms were reported in the largest randomised controlled trials.[153;156;159;161;162]

8.5 Summary

In summary, neonatal hypothermia appears to reduce neurodevelopmental disability in survivors of neonatal hypoxic-ischaemic encephalopathy, and is likely to be cost-effective. It has now become an established treatment within UK practice. However, some concerns exist that there is a potential for bias in the published trials, given that the nature of the treatment blinding is impossible, and that a clinician's decision to withdraw treatment to a neonate may be influenced by awareness of the treatment allocation and trial participation.[164] In order to monitor the ongoing effects of therapeutic hypothermia, within the UK all infants who undergo this treatment should be added to the TOBY Cooling Register.[165]

CHAPTER 9: ASSESSING THE IMPACT ON THE COST-EFFECTIVENESS OF NEONATAL HYPOTHERMIA OF VARYING THE THRESHOLD OF CORD PH FOR TREATMENT. A DECISION- ANALYTIC MODEL BASED ANALYSIS

9.1 Abstract

9.1.1 Background

Neonatal therapeutic hypothermia has been shown to be a cost-effective treatment in reducing the risk of cerebral palsy in infants with hypoxic ischaemic encephalopathy. A low umbilical cord pH at birth is one of the criteria used to define eligibility for this treatment. However, the threshold used to define a low pH is based on consensus statement rather than high quality evidence. The purpose of this analysis was to investigate the economic impact of varying the threshold of umbilical cord pH used to determine the need for neonatal cooling.

9.1.2 Methods

A decision-analytic model based cost-effectiveness analysis, based on data from systematic reviews. The analysis was performed from the UK NHS perspective and the main outcome was cost per case of cerebral palsy avoided.

9.1.3 Results

Using a cord pH threshold of <7.00 in combination with neonatal cooling dominated the higher threshold of <7.10. The incremental cost effectiveness ratio (ICER) for a cord pH threshold of <7.00 in combination with neonatal cooling in comparison to no test or treatment was £18733459 per case of cerebral palsy avoided. Sensitivity analyses varying the cerebral palsy prevalence rate, accuracy of cord pH testing and costs reduced the ICER but did not change the overall result.

9.1.4 Conclusion

When compared to a higher threshold of <7.10, the current cord pH threshold of <7.00 as one of the factors to indicate a neonate should receive therapeutic hypothermia is more effective and less costly and therefore the dominant strategy. Further research is needed to determine the cost-effectiveness of cord pH as a test in the general neonatal population.

9.2 Introduction

Cerebral palsy is a disorder affecting voluntary movement and posture, and is the commonest cause of motor impairment and significant physical disability in childhood.[2] Therapeutic neonatal hypothermia, as described in section 8.2.2, has been shown to improve neurologic outcomes in survivors of perinatal asphyxia,

compared to standard care.[159] An economic evaluation (Section 8.2.6) showed that cooling was likely to be cost effective.[163]

Neonatal cooling is now widely used within the UK to reduce the incidence of adverse outcome after a hypoxic insult. The eligibility criteria for treatment most commonly employed are those for the TOBY trial inclusion criteria, as described in section 8.2.2.[58;159]

Umbilical cord pH at birth is a measure of hypoxia, and a low cord pH has been associated with cerebral palsy. However, the threshold of 7.00 has been defined by consensus statement rather than clinical evidence, and the optimal cut-off for a low pH in predicting adverse outcome remains uncertain.[71] The systematic review reported in Chapter 5 collated primary studies, defining a low arterial pH at a variety of thresholds (7.00-7.20) and established a prognostic association for cerebral palsy. However, the data was insufficient to determine the pH threshold which should be used in clinical practice.[166]

This chapter reports the results of a model based cost-effectiveness analysis, which uses evidence from systematic reviews[166] (Chapter 5) on test accuracy of umbilical artery pH, and effectiveness[150] (meta-analysis reported in section 8.2.2) from the perspective of the NHS in the UK, to assess the impact of varying the pH threshold used to define eligibility for neonatal cooling. The primary outcome is based on the additional cost per case of cerebral palsy avoided.

9.3 Methods

9.3.1 Model structure

A decision tree model was used to perform an economic evaluation comparing different test- treatment strategies. A decision tree was felt to be the most appropriate model type for this analysis due to the short term nature of the decision problem:[167] cerebral palsy is a chronic condition, without possibility of cure, and the treatment is a single course, performed within 6 hours of delivery,[58] without scope for change in effectiveness over time. The tree was constructed in Data Treeage Pro 2012.[168]

The evaluation performed was a cost-effectiveness analysis based on the cost per case of cerebral palsy avoided. The analysis was performed from a UK health provider perspective, both community and hospital based costs were considered. A time horizon of 18 months was utilised on pragmatic grounds, based on the availability of cost data for this period.[163] Analyses from a societal perspective, which might attempt to include private costs to the individual, were considered beyond the scope of this evaluation, although it is acknowledged that the nature of the condition is likely to carry lifelong additional costs at an individual, health care and societal level.

Five different treatment strategies were compared:

1. No umbilical cord pH or cooling performed (‘no test/ no treatment’)
2. Umbilical cord pH performed (threshold considered positive <7.00), no treatment regardless of result. (‘cord pH <7.00/no treatment’)
3. Umbilical cord pH performed (threshold considered positive <7.00), neonatal cooling if positive. (‘cord pH <7.00/neonatal cooling_positive’)

4. Umbilical cord pH performed (threshold considered positive <7.10), no treatment regardless of result. (‘cord pH <7.10/no treatment’)
5. Umbilical cord pH performed (threshold considered positive <7.10), neonatal cooling if positive. (‘cord pH <7.10/neonatal cooling_positive’)

A subsection of the tree is presented in Figure 9.1. This shows branches 1-3. The other branches are analogous. In the figure, each branch to the right of the chance node (round symbol) indicates a possible outcome. The formula given beneath the branch (and in the box to the left of the root node (square symbol)) indicates how the probability of each outcome is calculated: for example when a test/ treatment combination is used, this considers the positive or negative predictive value of the test and the relative risk of the treatment for cerebral palsy. The terminal node (triangle symbol) indicates the end of the pathway and whether the condition is present or absent. The formula for calculating costs accrued through the test, treatment and the cost of the condition are given following the terminal node. The model inputs are described subsequently.

Pathway 1 (no test/ no treatment) is the comparator for pathways 2-5. Pathway 3 ((‘cord pH <7.00/neonatal cooling_positive’) represents current clinical practice; the threshold of <7.00 to define acidosis and therefore a significant hypoxic insult is the most commonly accepted. Pathway 4 and 5 represent comparisons with using a threshold of <7.10 to define acidosis to determine the impact on cost-effectiveness of this alternative threshold. Testing of other thresholds was not possible due to a lack of accuracy or effectiveness data or both.

9.3.2 Model inputs

Accuracy data

Full details of the systematic review to evaluate the association of umbilical cord pH with cerebral palsy are reported elsewhere (Chapter 5). A total of 51 articles were selected for final inclusion in the review, including seven which had data allowing the generation of 2 x 2 tables (true positive, false positive, false negative, true negative) for a low umbilical cord pH and cerebral palsy. All of these studies had a population solely consisting of pre-term infants.

For the purposes of this analysis, the searches were updated to October 2011 (the search strategy is given in Appendix 7). However no further studies with data relating umbilical cord pH and cerebral palsy were identified. There were no studies reporting the use of pH in combination with the Apgar score or HIE to predict cerebral palsy. Of the seven studies presented in the original review, two used a threshold of cord pH <7.00 (Socol et al 1994, Ingemarrson et al 1997) and two a threshold of <7.10 (Beeby et al 1994, Murphy et al 1995) 2 x 2 data from these four studies were re-analysed to calculate sensitivity, specificity and positive and negative likelihood ratios of umbilical arterial cord pH at birth to predict cerebral palsy. There was insufficient data to analyse other thresholds. Random effects meta-analysis was performed for each threshold (7.00 and 7.10); the data are presented in Table 9.1. Analyses were performed using Meta-Disc.[105] The pooled sensitivity and specificity results for each threshold were used, and converted to positive and negative predictive values within the model using the following formulae: Positive predictive value (PPV)= (sensitivity x prevalence)/((sensitivity x prevalence)+((1- specificity)x(1-prevalence))); Negative

predictive value (NPV)= (specificity x(1-prevalence))/(((1-sensitivity)x prevalence)+(specificity x(1-prevalence))).

Effectiveness data

A Cochrane review of the effectiveness of cooling for neonates with hypoxic ischaemic encephalopathy was identified, and the findings reported in section 8.2.1.[150] A subsequent meta-analysis published by Edwards et al (summarised in section 8.2.3) gives effectiveness data on the rate of cerebral palsy following neonatal hypothermia of the three trials using pH <7.00, with follow up to 18 months of age.[160] Data from this analysis regarding the relative risk and 95 % confidence intervals of cerebral palsy in survivors following hypothermia were included in the model; this data is reproduced in Table 9.2. The data from the two trials published following this meta-analysis (summarised in sections 8.2.4 and 8.2.5) were not included due to differences between the protocols used for treatment and UK practice.[161;162] Data regarding the relative risk of major neurodisability following hypothermia in the single trial using a threshold of pH <7.10[151] are also presented in Table 9.2, reproduced from the Cochrane review.[150] The primary study was obtained, but as unpublished data were provided to the Cochrane reviewers this data was felt to be the most complete and accurate.

Prevalence

A search of Medline and EMBASE was performed using the search terms ‘prevalence’ and ‘cerebral palsy’. A paper estimating the 2007 rate of newly diagnosed cases based on Surveillance of Cerebral Palsy in Europe (SCPE)[169] and Office of National Statistics (ONS) data regarding the prevalence and live birth rate

was identified, and felt to contain the most recent and relevant data.[2] A prevalence rate of 2 to 2.5 per 1000 live births was quoted. The lower estimate of 2 per 1000 was included for the base case analysis.

Costs

A search of Medline, EMBASE, the Cochrane Library and Web of Science was performed in October 2011. A combination of MeSH headings and keywords including ‘cerebral palsy’, ‘cost- benefit analysis’, and ‘neonatal cooling’ were used. The search strategy is given in Appendix 17. Only one existing UK based economic evaluation of the cost-effectiveness of therapeutic hypothermia was identified, the findings are summarised in section 8.2.6.[163]

The data presented in this paper represent the most recent and accurate data regarding the cost of neonatal hypothermia and the associated course. No alternative estimates of the cost of cerebral palsy was identified, therefore the data presented in this paper was used for this as well.[163] For the purposes of this model, the costs reported for the non-cooled group until 12 months of age were used for all infants in the non-treatment arms who had a low umbilical cord pH at either threshold. The cost for neurodevelopmental delay in the non-cooled group (from 12-18 months) was added for those who subsequently developed cerebral palsy. The same cost was used in the ‘no test, no treatment’ arm. In those who did not develop cerebral palsy, the costs for the non-cooled group without developmental delay (12-18months) were used. In the branches where those who had a low pH were treated with therapeutic hypothermia, the costs related to those undergoing treatment in the TOBY trial were

used. This allowed a timescale for the model and associated costs of 18 months. Details of all these costs are presented in Table 9.3.

All infants entering the TOBY trial had evidence of a hypoxic ischaemic insult, including a cord pH <7.00. As this model also includes neonates who had a normal pH at birth, and did not suffer hypoxia, the health care costs until 18 months of age of a neonate without hypoxic ischaemic encephalopathy, representing the general population, were also defined. No single cost defining this was identified from the literature, and therefore a variety of sources were used. The summary costs entered into the model are given in Table 9.3. The full calculations for costs of a non-encephalopathic infant are presented in Appendix 18, and summarised below.

At delivery, a healthy baby does not incur costs for delivery events separately from the mother.[170] Centre for Health Economics data estimate that, of the activity within obstetric departments, 0.1 % of obstetric patients would have a neonate with multiple minor diagnoses, and 0.7% would have a neonate with one minor diagnosis.[171] When restricted to delivery events only, this translated to 0.2% and 1.4% respectively. These were multiplied by the reported Healthcare Resource Group (HRG) costs, given in the same paper, to obtain a neonatal inpatient cost for the average obstetric delivery.[171]

According to Office of National Statistics (ONS) data for 2007/2008, the hospital admission rate in the 0-5 year age group was 128.9 per 1000.[172] This was used to calculate the likelihood of an admission for an individual during the first 18 months of life and multiplied by the cost of a bed day[163]. The frequency of GP consultations in the 0-4 month age group in 2008/2009 were identified from the literature.[173] These

were multiplied by reported costs of a 10 minute consultation.[174] The cost of child health surveillance to age 18 months was also obtained.[175] It was assumed that the average child would visit the practice nurse on four occasions for immunisations and these costs were included.[174]

The cost of the umbilical cord blood analysis for pH was obtained from Nottingham University Hospitals NHS trust. This was added to the cost of 10 minutes of midwifery time to perform the test.[174] All costs, where otherwise reported, were inflated to 2010/2011 costs using recommended methods.[174]

Figure 9.1 Diagram showing pathways 1-3 of the decision tree

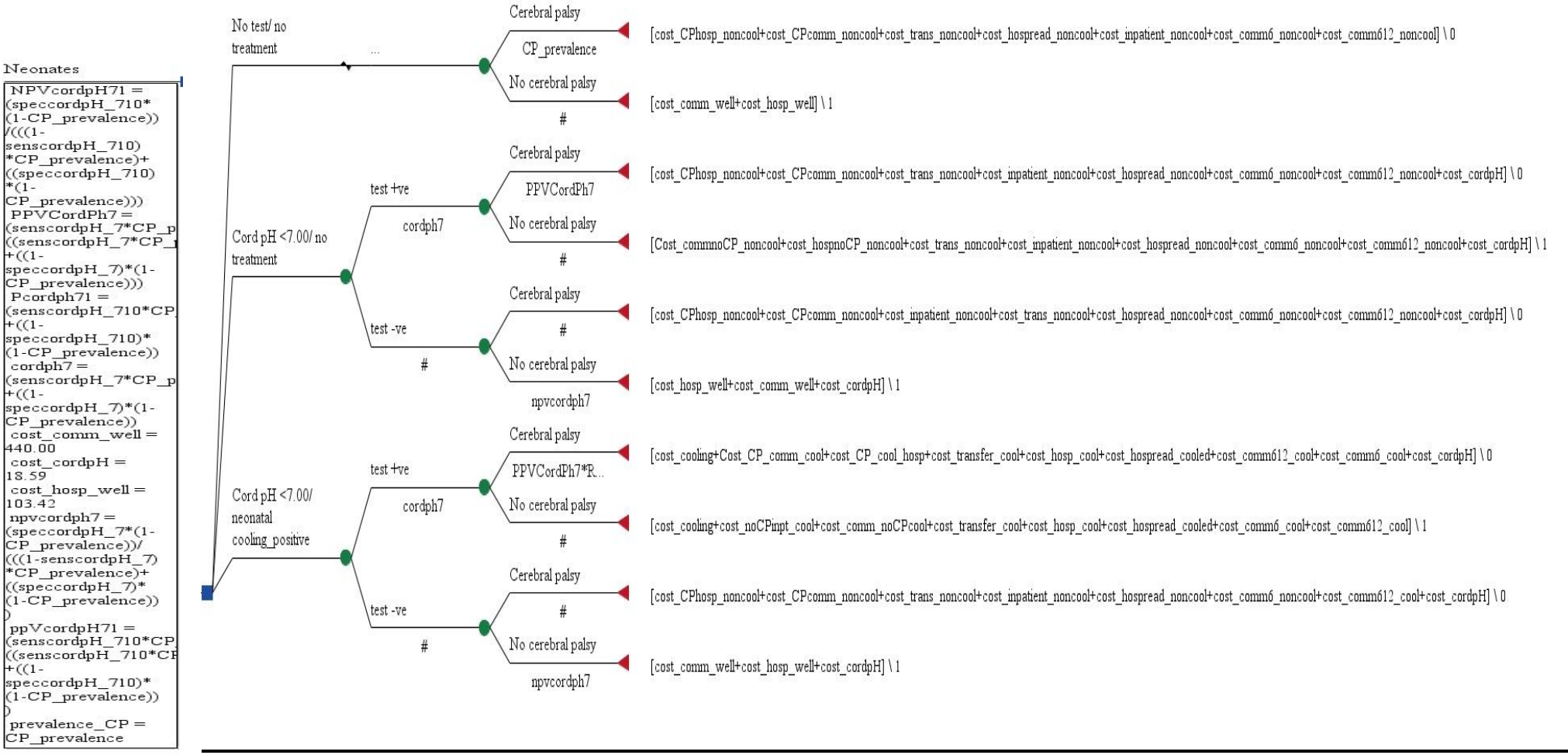


Table 9.1 Summary of the accuracy of umbilical arterial cord pH to predict cerebral palsy

Study	Cord pH threshold	Sensitivity (95% CI)	Specificity (95% CI)	Positive Likelihood ratio (95% CI)	Negative Likelihood ratio (95% CI)
Ingemarrson et al 1997	7.00	0 (0-0.84)	0.70 (0.63-0.76)	0.55 (0.04-6.90)	1.20 (0.72-2.01)
Socol 1994	7.00	0.83 (0.34-0.97)	0.50 (0.12-0.88)	1.67 (0.69-4.00)	0.33 (0.05-2.37)
Meta-analysis (2 studies)	**7.00**	**0.63 (0.25-0.92)**	**0.69 (0.62-0.75)**	**1.48 (0.65-3.38)**	**0.77 (0.16-3.83)**
Beeby et al 1994	7.10	0.05 (0.001-0.25)	0.93 (0.89-0.96)	0.73 (0.10-5.26)	1.02 (0.92-1.13)
Murphy et al 1995	7.10	0.25 (0.10-0.47)	0.88 (0.81 - 0.93)	2.00 (0.87-4.59)	0.86 (0.67-1.09)
Meta-analysis (2 studies)	**7.10**	**0.16 (0.07-0.30)**	**0.91 (0.88-0.94)**	**1.72 (0.80-3.70)**	**0.95 (0.77-1.19)**

Table 9.2 Summary of the effectiveness of neonatal therapeutic hypothermia versus standard care on neurodevelopmental outcome in survivors

Outcome	Number of primary studies	Source	Cord pH threshold for intervention	Relative risk (95 % CI)
Cerebral palsy in survivors	3	Edwards et al meta-analysis[160]	<7.00	0.69 (0.54 to 0.89)
Major neurodevelopmental disability in survivors	1	Cochrane review[150]	≤ 7.09	2.67 (0.35-20.51)

Table 9.3 Summary NHS costs per patient for tests, intervention and outcome

Cost Item	Cost (UK£ 2010/2011) (95% CI)[a]	Source
Test		
Umbilical cord pH (midwife to perform test (10mins) and run on point of care analyser)	18.59	1. Nottingham University Hospitals NHS Trust 2. Curtis[b]
Treatment cost		
Cost of cooling per infant including equipment and aEEG	6539 (1488-15806)	Regier et al[c]
Inpatient costs per cooled infant (survivors)	13747 (11416-16323)	Regier et al[c]
Transfer costs per cooled survivor	162 (136-193)	Regier et al[c]
Costs of hospital readmission per cooled survivor	1200 (694-1835)	Regier et al[c]
Community care costs to 6 months of age per cooled survivor	753 (614-902)	Regier et al[c]
Community care costs 6-12 months of age per cooled survivor	486 (336-1769)	Regier et al[c]
Inpatient costs per surviving encephalopathic infant (non-cooled)	14064 (11661-16664)	Regier et al[c]
Transfer costs per encephalopathic survivor (non-cooled)	159 (131-190)	Regier et al[c]
Cost of hospital readmission per encephalopathic infant (non-cooled)	2,649 (1260-4434)	Regier et al[c]
Community costs to 6 months of age per encephalopathic infant (non-cooled)	789 (610-991)	Regier et al[c]
Community costs 6-12 months of age per encephalopathic infant (non-cooled)	756 (482-1098)	Regier et al[c]
Community costs per non-encephalopathic infant to 18 months of age	440	1.Hippisley-Cox et al[d] 2.Curtis[b] 3.Sanderson et al[e]
Hospital costs per non-	120.82	1. Laudicella et al[f]

encephalopathic infant to 18 months of age		2. Regier et al[g] 3. ONS data[h]
Inpatient costs per case of cerebral palsy 12-18 months of age (cooled encephalopathic infants)	1744 (670-3309)	Regier et al[c]
Community costs per case of cerebral palsy 12-18 months of age (cooled encephalopathic infants)	373 (197-611)	Regier et al[c]
Inpatient costs per encephalopathic infant without cerebral palsy 12-18months of age(cooled)	137 (46-281)	Regier et al[c]
Community costs per encephalopathic infant without cerebral palsy (12-18 months of age (cooled)	236 (125-385)	Regier et al[c]
Inpatient costs per case of cerebral palsy 12-18 months of age (non-cooled encephalopathic infants)	1221 (411-2484)	Regier et al[c]
Community costs per case of cerebral palsy 12-18 months of age (non-cooled encephalopathic infants)	507 (287-783)	Regier et al[c]
Inpatient costs per encephalopathic infant without cerebral palsy 12-18months of age(non-cooled)	143 (32-338)	Regier et al[c]
Community costs per encephalopathic infant without cerebral palsy (12-18 months of age (non-cooled)	236 (125-385)	Regier et al[c]

a. All costs inflated to common price year of 2010/2011 as per payment and prices index.[150;174]
b. Cost of midwife, GP and practice nurse patient contact time.[174]
c. Costs are mean costs per cooled and non-cooled infant in the TOBY trial[163]
d. Average number of GP visits per year in the 0-4 year age group[173]
e. Cost of routine child health surveillance to 18 months (3 visits)[175]
f. Number and HRG cost of minor neonatal illnesses associated with delivery in obstetric unit[171]
g. Cost of a hospital readmission[163]
h. Average number of hospital admissions per year in 0-5 year age group[172]

9.3.3 Analysis

The main outcome of the model was cost per case of cerebral palsy avoided. The base case analysis was deterministic, in which the point estimates for test accuracy and treatment effectiveness were used to compare the cost effectiveness of the different pathways described above. Incremental cost effectiveness ratios (ICER) of the different strategies were calculated, and the results presented in a cost-effectiveness plane.

Deterministic sensitivity analysis 1: As the cost of a non-encephalopathic baby was estimated rather than obtained from an existing source, a sensitivity analysis was performed where this cost was doubled, and multiplied by 10, and 100, to assess the impact on the results of the true cost being higher than estimated.

Deterministic sensitivity analysis 2: The prevalence of cerebral palsy was increased to 2.5 per 1000 as this was the higher estimate obtained from the literature.

Deterministic sensitivity analysis 3: A sensitivity analysis was performed to determine if improving the accuracy of the cord pH test would change the relative cost effectiveness of the strategies examined. The sensitivity and specificity of cord pH <7.00 to predict cerebral palsy were increased. It was recognised that in reality sensitivity and specificity are related and there is a trade-off in that when sensitivity is increased, specificity is reduced. However, the purpose of this analysis was to explore the effect of a more accurate test overall, and therefore arbitrary figures of higher sensitivity were selected, without altering the specificity accordingly.

Probabilistic sensitivity analysis (PSA): The purpose of this analysis was to reflect the uncertainty surrounding the input parameters into the decision model, and examine the influence of uncertainty on the cost-effectiveness analysis. In order to perform this analysis, each of the model inputs (sensitivity, specificity, relative risks and, where 95% confidence intervals available, costs) were assigned a distribution. For the prevalence, a beta (β) distribution was used, with the parameters α = number of events of interest and β= total population- events of interest.[176] A β distribution was also used for sensitivity and specificity. In order to calculate α,β, an Excel spreadsheet was used to compare actual mean values, lower and upper 95% confidence intervals with fitted values to obtain the 'best fit' distribution. For relative risks, a log normal distribution was used, through taking the natural logs of the point estimate and 95% confidence intervals and using the following formula to calculate the estimate of log scale standard error:[176]

$$SE\,[\ln(RR)] = \frac{\ln(\text{upper 95\% confidence interval}) - \ln(\text{lower 95\% confidence interval})}{2 \times 1.96}$$

The log value of the point estimate and the SE[ln (RR)] were then used to define the distribution. For costs, a gamma distribution was assigned, again using an Excel spreadsheet to obtain α and λ values according to the best fit as described above. A definition of the variables and distributions used in the Treeage model are given in Appendix 19.

A Monte Carlo simulation with 10000 iterations was then performed to assess the effect of the uncertainty around parameters on the likely cost-effectiveness of any given test- treatment pairing.

9.4 Results

Base case: The results of the deterministic cost- effectiveness analysis are presented in Table 9.4. 'No test, no treatment' dominated all strategies except 'cord pH <7.00, cooling test positive'. The ICER for this strategy was £18733459 per case of cerebral palsy avoided. When 'no test, no treatment' was removed from the model, 'cord pH <7.00, cooling test positive' dominated the other strategies. The results are presented graphically on the cost-effectiveness plane in Figure 9.2. The nearer a strategy is to the bottom right corner of the graph, the greater the effectiveness, and cheaper the cost. A strategy will dominate others if it is both cheaper and more effective.

Deterministic sensitivity analysis 1: Table 9.5 gives the results of the deterministic analysis when the estimated costs of hospital and community care for a non-encephalopathic baby to age 18 months are increased. The ICER for 'cord pH <7.00/ neonatal cooling positive' is reduced with the increasing cost of a 'well' baby, but remains large even when the estimated cost is multiplied by 100 from the base case. All other strategies remained dominated.

Deterministic sensitivity analysis 2: A further sensitivity analysis was performed where the prevalence of cerebral palsy was increased to 2.5 per 1000, the higher figure quoted in the literature. This resulted in a reduction in the ICER for 'cord pH <7.00, neonatal cooling positive' to £14982150 per case of cerebral palsy avoided but did not alter the overall results.

Deterministic sensitivity analysis 3: Cord pH <7.00 does not have high accuracy to predict cerebral palsy, with the pooled sensitivity and specificity included in the

base case analysis being 0.63 and 0.69 respectively. Increasing the sensitivity of cord pH was to 0.80 led to both ‘cord pH <7.00, no treatment’ and ‘cord pH <7.10, no treatment’ being no longer dominated, and a reduction in the ICER for ‘cord pH <7.00, neonatal cooling positive’ to £2978551.32 per case of cerebral palsy avoided. However, increasing the sensitivity further to 0.9 led to a higher ICER of £13014159.

PSA: The results are presented in Figure 9.3 as a ‘cost-effectiveness acceptability curve’ (CEAC), which illustrates the probability that a given strategy is cost-effective according to the ‘willingness to pay’ threshold. It is not until a threshold of £25000000 that it becomes more likely that ‘cord pH <7.00, neonatal cooling positive’ is more cost-effective than ‘no test, no treatment’. At a threshold of £30000, the conventionally accepted cut-off of willingness to pay in the UK health service, there is a very low probability that neonatal cooling in combination with cord pH is cost effective.[177]

Table 9.4 Base case analysis results, costs, effectiveness and ICER for test/ treatment combinations for all neonates

Strategy	Cost (£ 2010/2011)	Incremental Cost	Effectiveness	Incremental Effectiveness	ICER[a](£ 2010/2011)	C/E	
Excluding dominated							
No test/ no treatment	599.99		1			601.19	
Cord pH <7.00/ neonatal cooling positive	7727.34	7127.35	1	0	18733459	7739.87	
All							
No test/ no treatment	599.99	0	1	0	0	601.19	
Cord pH <7.10/ no treatment	2220.1	1620.11	1	0	0	2224.55	Dominated
Cord pH < 7.10/ neonatal cooling positive	2796.27	2196.28	1	-0	-1926385	2805.08	Dominated
Cord pH <7.00/ no treatment	6278.49	5678.5	1	0	0	6291.07	Dominated
Cord pH <7.00/ neonatal cooling positive	7727.34	7127.35	1	0	18733459	7739.87	
All by Increasing effectiveness							
Cord pH < 7.10/ neonatal cooling positive	2796.27		1			2805.08	
No test/ no treatment	599.99		1			601.19	
Cord pH <7.00/ no treatment	6278.49		1			6291.07	
Cord pH <7.10/ no treatment	2220.1		1			2224.55	
Cord pH <7.00/ neonatal cooling positive	7727.34		1			7739.87	

Figure 9.2 Base case results for cost-effectiveness analysis

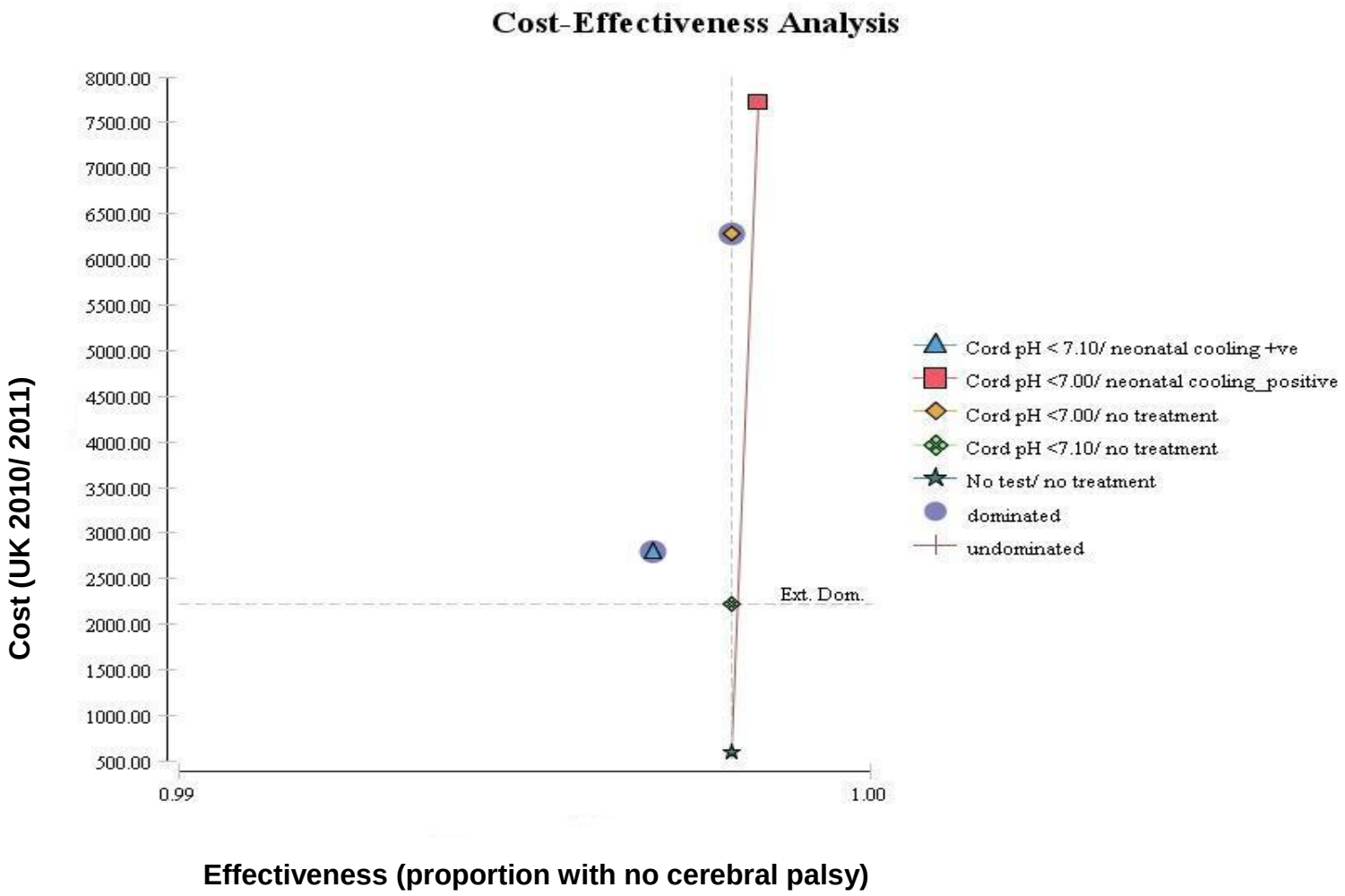


Figure 9.3 Results of probabilistic sensitivity analysis. Probability that different strategies are cost-effective at different 'willingness to pay' thresholds

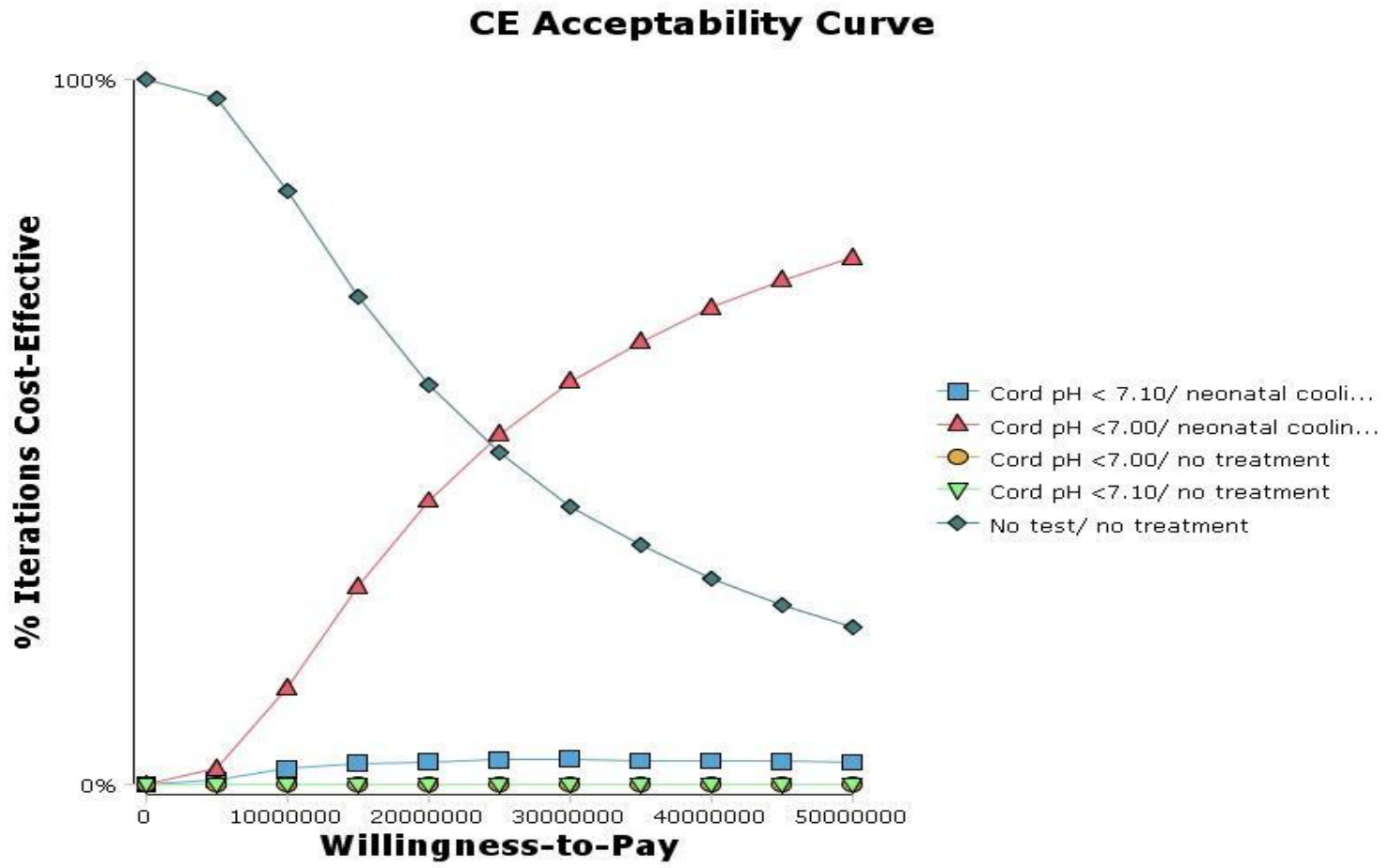


Table 9.5 Deterministic sensitivity analysis 1: Varying the cost of a care of a non-encephalopathic baby is varied from the base case cost of £ 560.82

Strategy	Cost (£ 2010/2011)	Incremental Cost	Effectiveness	Incremental Effectiveness	ICER (2010/2011)	C/E
Base cost doubled (£1121.64)						
No test/ no treatment	1159.69		1			1162.01
Cord pH <7.00/ neonatal cooling_positive	8112.97	6953.28	1	0	18275945	8126.13
Base cost x 10 (£5608.20)						
No test/ no treatment	5637.27		1			5648.57
Cord pH <7.00/ neonatal cooling_positive	11198.03	5560.75	1	0	14615833	11216.19
Base cost x 100 (£56,082)						
Cord pH <7.00/ no treatment	44456.08		1			44545.17
Cord pH <7.00/ neonatal cooling_positive	45904.92	1448.85	1	0	3808136	45979.39

9.5 Discussion

The main finding is that in all analyses using a cord pH threshold of <7.00 and treating those who were test positive with neonatal cooling dominated the strategy of using a cord pH threshold of < 7.10 to determine treatment. The results support current clinical practice, where a threshold of <7.00 is used in conjunction with other criteria to identify neonates to undergo neonatal therapeutic hypothermia, and suggest that raising the threshold of pH used would be a less effective (in terms of cases of cerebral palsy avoided) and more costly strategy.

Strengths of the economic evaluation

The strengths of this evaluation lie in the methodology used. The author received training in decision-analytic modelling and advice from experts in health economic evaluation (Professor Tracy Roberts) and regarding the clinical assumptions (Dr Helen Budge, Dr Andy Ewer, Dr Katie Morris). The data used to inform the model were based upon high quality systematic reviews with recently updated searches and was felt to be as complete as possible. A thorough literature search was performed to obtain the most recent cost data, and where none were available to calculate the cost in a thorough manner considering all major community and hospital factors.

Limitations of the economic evaluation

There are several limitations to the analysis. A number of assumptions had to be made due to the availability of data and scope of the evaluation, which may have affected the results. These are set out in Table 9.6.

Table 9.6 Limitations of the economic evaluation

Area of evaluation	Details of limitation	Implications
Accuracy data		
1.Quality of primary studies	The studies on which the accuracy data was provided are small and of low quality	The reliability of the accuracy data may be affected
2. Population of primary studies	The infants included in the primary studies were all preterm, and the population in which neonatal hypothermia is performed is greater than 36 weeks gestation	The accuracy in this population may differ from that in the target population
3. Accuracy data available for cord pH alone	The criteria for a neonate to be treated with hypothermia are not based on the cord pH result alone, but also other evidence of a hypoxic ischaemic encephalopathy including low Apgar score, and abnormal neurological signs. No data were available for different combinations of these tests and signs so these were not included in the model	The exclusion of other factors which would affect the decision to cool may be valid for a comparison between the strategies of treating at cord pH threshold <7.00 and <7.10 but this cannot be proven.

4. Differences in tests performed in clinical trials of treatment effectiveness	The criteria for cooling are not only based on an umbilical cord pH at birth below the specified threshold, but also a pH sample from the baby's arterial blood up to an hour after delivery. Accuracy data from the systematic reviews focused on cord pH only.	The accuracy data from the systematic reviews regarding pH may differ from the entry criteria in the trials providing the effectiveness data.
5. Only 2 thresholds of pH could be analysed	There were only sufficient accuracy and effectiveness data available to explore a pH threshold of <7.00 and <7.10	Other thresholds could not be compared to the current standard of pH <7.00
Effectiveness data		
1.Quality of primary studies	The three studies which used a pH threshold of <7.00 were well powered, high quality randomised controlled trials. The single study using a threshold of <7.10 was smaller and of lower quality	The effectiveness data for a neonatal hypothermia at a pH threshold of <7.10 is likely to be a less reliable estimate
2. Outcome defined in primary studies	For the three studies with a threshold of <7.00, data was available regarding the specific outcome of cerebral palsy. However, for the study with a threshold of <7.10, the outcome was neurodevelopmental delay, which may include other diagnoses.	The effectiveness data for the <7.10 study may differ from the diagnosis in question. However, for the other three studies (threshold <7.00) data on neurodevelopmental delay were also available and the effectiveness (relative risk) was similar to that of cerebral palsy so this assumption is likely to be valid
Costs		

1.Cost of cerebral palsy	The only cost data available in the literature was obtained from a cost-effectiveness analysis run alongside a trial of neonatal cooling. The cost data specified was of neurodevelopmental delay to age 18 months, diagnosed at age 12 months	The costs of cerebral palsy were assumed to be the same for neurodevelopmental delay, but this may not be the case.
2. Medicolegal costs are not included	Cerebral palsy resulting from a hypoxic insult suffered during the perinatal period is a significant source of medicolegal claims and may result in large sums being paid in compensation	It was felt that trying to calculate the number and amount of successful medicolegal claims was beyond the scope of this analysis. The true cost of cerebral palsy including these is likely to be higher, reducing the ICER of a strategy including neonatal cooling.
3. Cost of a non-encephalopathic baby	These costs were calculated from a variety of sources, based on data regarding the average number of GP visits and hospital admissions for the general population in the 0-5 age group.	The data for the 0-18 month period, the time horizon of the model, may differ from this and the cost estimations may be inaccurate. However, this was addressed through sensitivity analysis by increasing the cost.
Model structure and assumptions		

1.Time horizon	The time horizon in the model is 18 months. However, cerebral palsy is an incurable condition with implications for lifelong health.	The costs of cerebral palsy with a longer time horizon are likely to be significantly higher and therefore it is likely that the true ICER of a strategy including neonatal hypothermia is much less than estimated in this model
2. Babies who have a normal pH at birth assumed to be non-encephalopathic	A ‘test negative’ baby with a normal pH at birth who did not develop cerebral palsy was assumed to be non-encephalopathic without any other specific health concerns. Babies who had a low pH were assumed to be encephalopathic and assigned a higher cost, however in reality cord pH is not a perfect test for encephalopathy and many babies with a low pH will be otherwise well, not requiring further intervention or special care, and conversely babies with a normal pH may have other problems	Attempting to define the number of babies with a normal pH who had encephalopathy and those with a low pH who were normal and include this in the model was felt to be beyond the scope of this evaluation. However, this is a significant deviation from the clinical reality and therefore likely to affect the true ICER.

3. Babies who did not go on to develop cerebral palsy in the ‘no test, no treatment’ strategy were assumed not to have had neonatal encephalopathy in the costs assigned	Of the babies who have neonatal encephalopathy, only a small proportion develop cerebral palsy, and therefore the cost of those without cerebral palsy is likely to be higher as some will have had neonatal encephalopathy and required a higher level of hospital and community care than assigned in the model.	Attempting to calculate the proportion of babies who did not have cerebral palsy but had neonatal encephalopathy or another significant diagnosis requiring extra care was felt to be beyond the scope of this evaluation, the main purpose of which was to compare the two thresholds of cord pH and neonatal cooling. Attempts were made to address this by increasing the cost assigned to a non-encephalopathic infant in the sensitivity analysis.

Where possible, attempts were made to address these limitations. The systematic review searches for accuracy data were updated, however no further studies were obtained and no data were available relating to a population born at 36 weeks gestation or greater. Similarly, no data were identified relating to the combined accuracy of cord pH with other factors, such as Apgar score or neurological status in the prediction of cerebral palsy, and it was therefore felt that attempting to factor these into the model would be overly complex and rely on too many assumptions.

The potential inaccuracy of the mean estimates of sensitivity, specificity and relative risks were recognised, however as these were the best available according to the current literature, probabilistic sensitivity analysis was performed, allowing sampling from distributions around the point estimate derived from the 95% confidence intervals to explore this uncertainty. This did not significantly affect the results. The differences in the population from which the accuracy data was derived may mean that the true accuracy in the target population is different, and the analysis where the estimated sensitivity was varied was performed to address this (deterministic sensitivity analysis 3). A higher sensitivity changed the magnitude of the results but did not affect the ranking according to cost effectiveness of the strategies examined.

In this evaluation, a strategy of not performing a cord pH test or treating with hypothermia appeared more cost-effective than testing and treating those with a cord pH <7.00 with neonatal cooling to prevent cerebral palsy, with the ICERs in all analyses being large and well above current ‘willingness to pay’ thresholds in the setting of the UK health service. The sensitivity analyses performed show

that this may be partly explained by the cost assumptions made, as raising the cost of a non-encephalopathic baby reduced the ICER, and it is likely that the true cost of these infants is higher than estimated. However, none of the analyses brought the ICER down to £20000- £30000 or less, which is the typical threshold used to approve a treatment as cost-effective.[177]

In comparison with the results of the analysis by Regier et al, they found that the incremental cost per disability free life year (DFLY) gained was £19931, with a 69% probability that cooling is cost effective at a willingness to pay threshold of £30000, also stating that neonatal hypothermia is likely to be cost-effective if a longer time horizon is considered.[163] The differences in results compared to their analysis is likely to be due to the fact that in their analysis all infants included were encephalopathic, therefore providing a homogeneous population in which to compare the cooling treatment, and test results were not considered.

Despite the limitations of this evaluation, the main finding, that a cord pH of threshold <7.00 is more appropriate than a threshold of < 7.10 as a criterion for performing neonatal hypothermia, is likely to be valid, as the assumptions apply to both branches of the model. Inferences regarding the actual cost effectiveness of this test treatment strategy cannot be made from this model, due to the fact that the ICER obtained is likely to be inaccurate as a result of the limitations described in table 9.6. Several important questions regarding cord pH testing remain, such as the use of other thresholds of pH; whether cord pH testing should be performed on all neonates at birth; and the implications of an abnormal test in combination with the presence or absence of other features such as a low Apgar score and abnormal neurological signs.

Recommendations for research

A primary study such as a large cohort or population study whereby neonatal tests and observations are prospectively recorded, and long term follow up data obtained to obtain neurodevelopmental outcomes would provide the ideal dataset to examine the cost-effectiveness of performing neonatal tests, including cord pH. The outcome data of such a study could be used in conjunction with individual patient decision modelling to answer a number of outstanding questions, such as the optimum threshold and combination of neonatal tests, including neurological status, Apgar score and cord pH, to define the population in which neonatal cooling is most cost-effective. It could also be used to determine the cost-effectiveness of cord pH testing for all neonates, or whether the current strategy of just testing those believed to be at higher risk of a hypoxic episode is valid. Standardised reporting of smaller studies, including compliance with the STARD checklist, to facilitate IPD meta-analysis, may present a more feasible alternative to a single large study.

9.6 Conclusion

The conclusion of this health economic evaluation is that, when compared to a higher threshold of <7.10, the current cord pH threshold of <7.00 as one of the factors to indicate a neonate should receive therapeutic hypothermia is more effective and less costly than the alternative high threshold and therefore the dominant strategy. Further research is needed to determine the cost-effectiveness of cord pH as a test in the general neonatal population.

CHAPTER 10: CONCLUSION

10.1 Introduction

This thesis evaluated the prognostic and predictive ability of neonatal tests for short and long term outcomes. It explored the effect of varying the threshold of umbilical cord pH on the cost-effectiveness of neonatal therapeutic hypothermia. The objectives were achieved in that it reports:

1. Summary estimates for the prognostic association and, where possible, predictive ability of the following tests for adverse outcomes: umbilical cord pH and base excess at birth, all current standards for defining low birth weight, and the Apgar score.
2. A decision-analytic model based evaluation comparing the use of umbilical cord pH at thresholds of 7.00 and 7.10 combined with neonatal therapeutic hypothermia for acidotic infants.

Each of the previous chapters in this thesis includes a detailed discussion of the main findings and conclusions, considering any limitations. This chapter focuses on the main findings of the work and discusses general strengths and limitations, leading to recommendations for clinical practice and future research.

10.2 Summary of main findings

10.2.1 Summary of reviews of prognostic and predictive ability

- In total, 2383 manuscripts were read in full. 218 papers were included, 51 in the review of umbilical cord pH, 92 in the birth weight standards review and 87 in the Apgar score review. The total number of individuals

included per review were 479022 (cord pH and base excess), 3690080 (Apgar) and 23051541 (birth weight standards). The majority of the included studies were of retrospective cohort design, and most were of high or moderate quality.

- All of the neonatal tests assessed had a strong association with neonatal mortality, however even where the association was strong, the sensitivity and negative likelihood ratios were generally poor, indicating that a negative test does not change the odds of an adverse outcome.
- The association between the tests and long term outcomes varied.

10.2.2 Summary of decision-analytic modelling

- When compared to a higher threshold of <7.10, the pH threshold of <7.00 used in current practice as one of the factors to indicate a neonate should receive therapeutic hypothermia is more effective and less costly and therefore the dominant strategy. Comparison with other thresholds was not possible due to a lack of accuracy and effectiveness data.

10.3 Strengths of the thesis

To the best of the author's knowledge there have been no other systematic assessments of the prognostic and predictive ability of neonatal tests for short and long term adverse outcomes. The only other systematic review identified in this area considered the association of low umbilical cord pH and low Apgar score with neonatal mortality and cerebral palsy, but did not consider the predictive ability or address the effects of confounding factors as attempted in this thesis.[77] No other reviews have compared different standards of low birth

weight and their association with adverse outcomes throughout the life course. It thus allows a very comprehensive evaluation of the knowledge to date, and with consideration of the strengths of limitations, recommendations for clinical practice and future research.

10.4 Limitations of the thesis

10.4.1 Reviews of prognostic and predictive ability

The limitations in the systematic reviews were generally related to:

- Reporting quality of the primary studies. In particular, poor reporting of the population characteristics, or failure to report results according to population subgroups, limited the analyses that could be done for different risk groups within a population, and therefore the extent to which the results can be extrapolated to clinical practice.
- The review methods: the fact that the tests were assessed in isolation and that individual patient data meta-analysis was not performed. However, given the age of some of the included studies and the overall poor response from authors when asked to provide data, the potential to acquire sufficient data for IPD or to assess tests in combination may be limited at this time.

It is felt however, that the robust methods used within the reviews addressed the limitations as far as possible, and that these reviews still represent the most up to date evidence synthesis for the tests investigated, and the use of advanced statistical techniques including the generation of prediction intervals

ensures the validity of the results including the likelihood of future studies in this area to generate similar point estimates.

10.4.2 Decision- analytic modelling

The limitations of the economic evaluation relate to both the model design and the primary data used to inform the model.

The design of the model was limited by:

- In order to avoid the model becoming over complex, and with the data available, many assumptions had to be made, including the fact that in practice other factors including the Apgar score and neurological signs and symptoms are taken into account in the decision to treat with hypothermia, and in the model the assumption was made that all other elements did not vary, and a baby with a normal pH was neurologically normal, which may not be the case in reality.
- The time horizon was limited to 18 months, from the NHS perspective, where cerebral palsy is a life-long condition, with many costs from a societal or individual perspective that are not be accounted for within the scope of the model.

The primary data was limited by:

- Due to the available accuracy and effectiveness data, only two thresholds of umbilical cord pH could be assessed.
- The population in which the accuracy data was obtained consisted of pre-term infants, whereas neonatal hypothermia is used in neonates of 36 weeks gestation or greater.

These limitations affect the overall cost-effectiveness estimate of cooling, which in this model is well beyond thresholds that would be funded by UK governing bodies. However, as the limitations apply to both thresholds examined, the finding that a cord pH threshold of 7.00, as used in current practice, dominates the use of a higher threshold of 7.10, is valid.

10.5 Recommendations for practice

The reviews in this thesis have shown that umbilical cord pH at birth, low birth weight and a low Apgar score have significant associations with a variety of adverse outcomes and are therefore important tests to consider. However, given the fact that in most cases a negative test was not found to change the likelihood of an adverse outcome, and the inability to fully assess the performance of tests in population subgroups or compare multiple tests in the same population, there are many unanswered questions regarding their use and interpretation in clinical practice. The main recommendations of this thesis are thus for future research, to clarify these issues.

10.6 Recommendations for research

Measurement of birth weight and recording the Apgar score are tests that are routinely performed after delivery. Further research, either in the form of individual patient data meta-analysis, or a large cohort study with long term follow up, are necessary to determine which definition of low birth weight most accurately predicts adverse outcome. Absolute birth weight by any threshold appeared to be more strongly associated with neonatal mortality than any other measure, however the different standards need to be compared in the same

population, and birth weight assessed as a continuous measure, to identify the optimum standard and threshold that can be used. Important subgroups, including different ethnic origins and body mass index, could also be considered by either of these methods. Again, more detailed subgroup analysis is necessary to identify the threshold of Apgar score which can be used to predict adverse outcome.

Umbilical cord pH at birth is not a test that is currently recommended for routine use. At present, targeted testing of infants deemed to be at high risk of asphyxia is performed. However, the findings of the systematic review performed in this thesis suggest that the associations with adverse outcome are stronger in an unrestricted population, suggesting that the cost effectiveness of performing the tests for all infants should be assessed. Additionally, further work assessing the pH as a continuous measure and assessing the result in combination with other tests or clinical findings are likely to improve the ability to predict adverse outcome through neonatal tests, facilitating counselling or targeting interventions.

A large cohort study or individual patient data meta-analysis of existing cohorts, such as population registry data, are likely to be the most effective methods of addressing the outstanding questions raised by this thesis. Standardised reporting of future studies, with adherence to the STARD checklist, would facilitate meta-analysis and would be an alternative to a single large cohort. High quality primary data could then be combined with effectiveness data in individual patient decision analytic modelling to determine the cost-effectiveness

of different test treatment strategies, including combinations of tests, to inform NHS practice.

APPENDICES AND REFERENCES

Appendix 1. Scoping search strategy

MEDLINE (Database inception – December 2006)

#	Search History	Results	Display
1	exp Infant, Newborn/	359997	DISPLAY
2	neonate.mp.	13149	DISPLAY
3	newborn.mp.	441692	DISPLAY
4	exp Birth Weight/	24194	DISPLAY
5	exp ANTHROPOMETRY/	63612	DISPLAY
6	exp Skinfold Thickness/	4431	DISPLAY
7	anthropometry.mp.	21342	DISPLAY
8	birthweight.mp.	7539	DISPLAY
9	birth weight.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	43797	DISPLAY
10	(birth weight adj gestation$).mp. [mp=title, original title, abstract, name of substance word, subject heading word]	1395	DISPLAY
11	(birthweight adj gestation$).mp. [mp=title, original title, abstract, name of substance word, subject heading word]	496	DISPLAY
12	(birthweight adj ratio$).mp. [mp=title, original title, abstract, name of substance word, subject heading word]	49	DISPLAY
13	(birth weight adj ratio$).mp. [mp=title, original title, abstract, name of substance word, subject heading word]	70	DISPLAY
14	(customised adj birth weight).mp. [mp=title, original title, abstract, name of substance word, subject heading word]	1	DISPLAY
15	(customised adj birthweight).mp. [mp=title,	6	

Appendix 1. Scoping search strategy

	original title, abstract, name of substance word, subject heading word]		DISPLAY
16	neonatal ponderal index.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	20	DISPLAY
17	mid-arm circumference.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	273	DISPLAY
18	MAC.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	8725	DISPLAY
19	skinfold thickness.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	5387	DISPLAY
20	weight for length.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	1880	DISPLAY
21	1 or 2 or 3	444090	DISPLAY
22	4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20	120997	DISPLAY
23	21 and 22	40179	DISPLAY
24	limit 23 to humans	38999	DISPLAY
25	limit 23 to animals	2022	DISPLAY
26	25 not 23	0	-
27	exp Cohort Studies/	547114	DISPLAY
28	exp MORTALITY/	158377	DISPLAY
29	exp MORBIDITY/	183577	DISPLAY
30	natural history.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	21030	DISPLAY
31	prognos$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	313190	DISPLAY

Appendix 1. Scoping search strategy

32	course.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	260108	DISPLAY
33	predict$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	420338	DISPLAY
34	exp "Outcome Assessment (Health Care)"/	263287	DISPLAY
35	outcome$1.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	550128	DISPLAY
36	(inception adj cohort$1).mp. [mp=title, original title, abstract, name of substance word, subject heading word]	679	DISPLAY
37	disease progression.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	47608	DISPLAY
38	exp Survival Analysis/	64631	DISPLAY
39	exp PROGNOSIS/	472662	DISPLAY
40	27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39	1887895	DISPLAY
41	23 and 40	19664	DISPLAY
42	(sensitivity and specificity$).mp. [mp=title, original title, abstract, name of substance word, subject heading word]	181215	DISPLAY
43	exp "Sensitivity and Specificity"/	193291	DISPLAY
44	sensitivity.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	420307	DISPLAY
45	specificity.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	503082	DISPLAY
46	((pre-test or pretest) adj probability).mp. [mp=title, original title, abstract, name of substance word, subject heading word]	490	DISPLAY
47	post-test probability.mp. [mp=title, original	148	

Appendix 1. Scoping search strategy

	title, abstract, name of substance word, subject heading word]		DISPLAY
48	predictive value$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	83781	DISPLAY
49	likelihood ratio$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	3044	DISPLAY
50	likelihood functions.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	7810	DISPLAY
51	exp Diagnostic Errors/	61428	DISPLAY
52	area under curve.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	16886	DISPLAY
53	((single or double or triple) adj blind$3).mp. [mp=title, original title, abstract, name of substance word, subject heading word]	111293	DISPLAY
54	reproducibility of results.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	114540	DISPLAY
55	double blind method/ or single blind method/	92858	DISPLAY
56	(randomized controlled trial or controlled clinical trial).mp. [mp=title, original title, abstract, name of substance word, subject heading word]	13043	DISPLAY
57	practice guideline.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	971	DISPLAY
58	consensus development conference$.mp. [mp=title, original title, abstract, name of substance word, subject heading word]	1628	DISPLAY
59	42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58	1050543	DISPLAY
60	41 and 59	1968	DISPLAY

Appendix 1. Scoping search strategy

The search strategy was adapted for use in Embase (2125 citations retrieved) and Cinahl (170 citations retrieved). A total of 3698 citations were available when duplicates were excluded.

Appendix 2. The STARD (Standards of Reporting for Diagnostic accuracy studies) checklist

Section and Topic	Item #		On page #
TITLE/ABSTRACT/ KEYWORDS	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	
INTRODUCTION	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	
METHODS			
Participants	3	The study population: The inclusion and exclusion criteria, setting and locations where data were collected.	
	4	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	
	5	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in item 3 and 4? If not, specify how participants were further selected.	
	6	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	
Test methods	7	The reference standard and its rationale.	
	8	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.	
	9	Definition of and rationale for the units, cut-offs and/or categories of the results of the index tests and the reference standard.	
	10	The number, training and expertise of the persons executing and reading the index tests and the reference standard.	
	11	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.	
Statistical methods	12	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	
	13	Methods for calculating test reproducibility, if done.	
RESULTS			

Appendix 2. The STARD (Standards of Reporting for Diagnostic accuracy studies) checklist

Participants	14	When study was performed, including beginning and end dates of recruitment.	
	15	Clinical and demographic characteristics of the study population (at least information on age, gender, spectrum of presenting symptoms).	
	16	The number of participants satisfying the criteria for inclusion who did or did not undergo the index tests and/or the reference standard; describe why participants failed to undergo either test (a flow diagram is strongly recommended).	
Test results	17	Time-interval between the index tests and the reference standard, and any treatment administered in between.	
	18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	
	19	A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	
	20	Any adverse events from performing the index tests or the reference standard.	
Estimates	21	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	
	22	How indeterminate results, missing data and outliers of the index tests were handled.	
	23	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	
	24	Estimates of test reproducibility, if done.	
DISCUSSION	25	Discuss the clinical applicability of the study findings.	

Appendix 3. The QUADAS (Quality Assessment of Diagnostic Accuracy Studies) checklist

Quality item	Response
Was the spectrum of participants representative of the patients who will receive the test in practice?	Yes/ No/ Unclear/
Were selection criteria clearly described?	Yes/ No/ Unclear/
Was the reference standard likely to classify the target condition correctly?	Yes/ No/ Unclear/
Was the period between performance of the reference standard and the index test short enough to be reasonably sure that the target condition did not change between the two tests?	For these reviews the answer will always be Not applicable
Did the whole sample or a random selection of the sample receive verification using the reference standard?	Yes (if ≥ 90%)/ No/ Unclear
Did participants receive the same reference standard regardless of the index test result?	Yes/ No/ Unclear/
Was the reference standard independent of the index test? (that is, the index test did not form part of the reference standard)	Yes/ No/ Unclear/
Was the execution of the index test described in sufficient detail to permit its replication?	Yes/ No/ Unclear/
Was the execution of the reference standard described in sufficient detail to permit its replication?	Yes/ No/ Unclear/
Were the index test results interpreted without knowledge of the results of the reference standard?	For these reviews the answer will always be Yes
Were the reference standard results interpreted without knowledge of the results of the index test?	Yes/ No/ Unclear/
Were the same clinical data available when the test results were interpreted as would be available when the test is used in practice?	Yes/ No/ Unclear/
Were uninterpretable, indeterminate or intermediate test results reported?	Yes/ No/ Unclear/
Were withdrawals from the study explained?	Yes/ No/ Unclear/Not applicable
Were interventions between the index test and reference standard reported?	Yes/ No/ Unclear/Not applicable

Appendix 4. Data extraction form for review of cord pH and base excess and adverse outcome

Data Collection Sheet Cord pH

Date: _____

Reviewer ID:_____ **Paper No:**_____

Year Of Publication:_____ **Language:**_____

Region study performed:_____

Section A: Study Selection

Population:

Inclusion: Children or adults who have had cord pH performed at birth

Yes ☐ (Include) No ☐ (study excluded)

Index test:

Cord pH performed:

Yes ☐ (include) No ☐ (study excluded)

Reference Test / Outcome Measure:

A reference standard looking at childhood or adult morbidity is performed

Yes ☐ (include) No ☐ (study excluded)

2 x 2 Table Possible:

Yes ☐ (include) No ☐ (exclude and case series ≤ 10, if other data e.g. sens, spec can include)

Study Selected:

Yes ☐ (must answer yes to ALL include questions)

No ☐ Give reason if NO _____

1)Ref ID:		**4)Publication year:**	
2)Rev		**5)First Author:**	

Appendix 4. Data extraction form for review of cord pH and base excess and adverse outcome

name:			
3)Country:		**6)Language:**	

Section B: Data Retrieval for Cord pH Study

Population

7) Healthcare Centre:

Secondary care ☐ Tertiary care ☐ Mixed ☐ Other ☐ Unreported ☐

8) Setting:

In-patient ☐ Unreported ☐ Other ☐_____

9) Number of participating centres: _____

10) Gestation at time of delivery:

20-24 weeks ☐ 24-28 weeks ☐ 28-34 weeks ☐ 34-37 weeks ☐ 37-40 weeks ☐

>40 weeks ☐ Unreported ☐ Other _____

10.i) Mean (range)_____ Unreported ☐

10.ii) Median (range) _____ Unreported ☐

11) Pregnancy:

Low Risk ☐ High Risk ☐ Unselected ☐ Unreported ☐

11.i) State high risk conditions: Unreported

12) Did all mothers have singleton pregnancies?:

Yes ☐ No ☐ Unreported ☐

13) Were mothers primigravid?:

Yes ☐ No ☐ Unreported ☐

14) What was the mode of delivery?

Appendix 4. Data extraction form for review of cord pH and base excess and adverse outcome

SVD ☐ Instrumental ☐ Emergency Caesarean section ☐

Elective Caesarean section ☐ Mixed ☐ Unreported ☐

15) Birthweight: < 3rd centile ☐ < 5th centile ☐ < 10th centile ☐ < 25th centile ☐

> 2SD ☐

<500g ☐ 500-1000g ☐ 1001-1500g ☐ 1500-2000g ☐ 2500-4000g ☐ >4000g ☐

Other ☐ _____ Unclear ☐

Unreported ☐

16) List other eligibility/ in-/exclusion criteria:

17) Study population: (describe age (mean +/- SD or median/range), ethnicity, social class, breastfeeding etc.

18) Start of patient inclusion (year) : _____ Unreported ☐

Appendix 4. Data extraction form for review of cord pH and base excess and adverse outcome

19) End of patient inclusion (year) : Unreported ☐

20) Study Design:

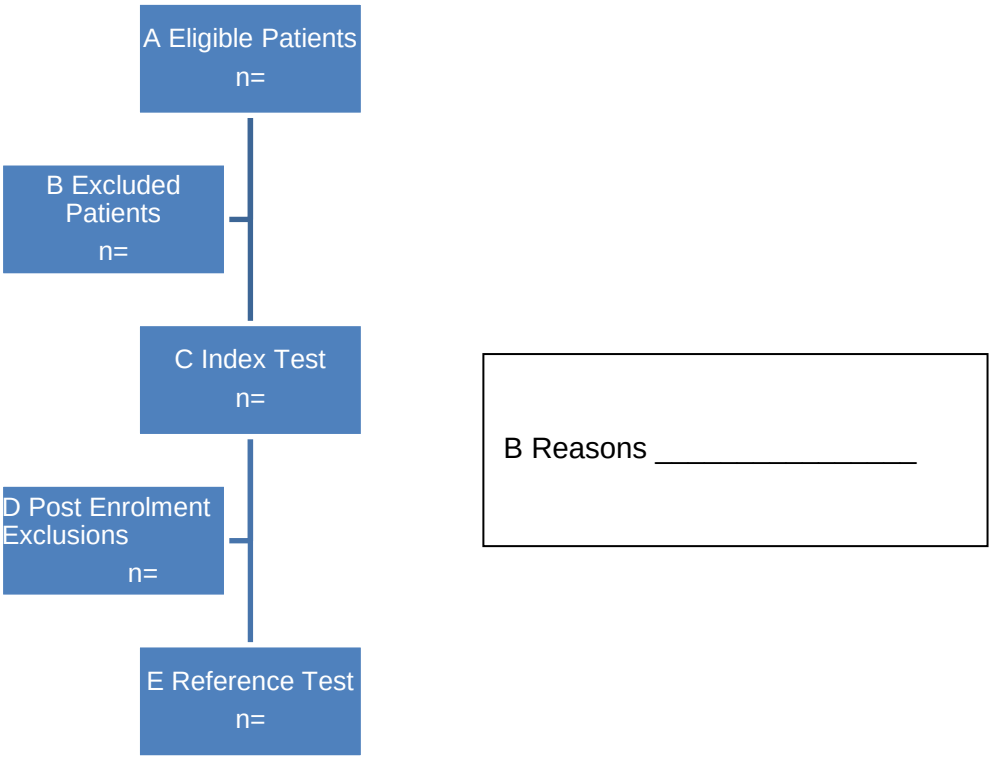
cohort ☐ case control ☐ RCT/CCT ☐ cross sectional ☐ before and after ☐

case series ☐ (no _____) other ☐

20.i) Data collection: prospective ☐ retrospective ☐ unreported ☐ other ☐

20.ii) Enrolment: consecutive ☐ arbitrary (random) ☐ unreported ☐ other ☐

21) Numbers:



21) Completeness of Verification:

(= E / C x 100 = %)

> 90% ☐ 81-90% ☐ < 81% ☐

Appendix 4. Data extraction form for review of cord pH and base excess and adverse outcome

Index Test

22) Description of technique:

Adequate ☐ Inadequate ☐

Cord pH:

23) Arterial: <6.90 ☐ 6.90-7.00 ☐ 7.01-7.10 ☐ 7.11-7.20 ☐ 7.21-7.30 ☐

7.31-7.35 ☐ >7.35 ☐

Median_____ Mean _____ Unreported ☐

Dataset used to establish threshold_____

Base excess: <2.0 ☐ 2.0-6.0 ☐ 6.1-10.0 ☐ 10.1-12.0 ☐ 12.1-16 ☐ >16 ☐

Median _____ Mean_____ Unreported ☐

24) Venous: <6.90 ☐ 6.90-7.00 ☐ 7.01-7.10 ☐ 7.11-7.20 ☐ 7.21-7.30 ☐

7.31-7.35 ☐ >7.35 ☐

Median_____ Mean _____ Unreported ☐

Dataset used to establish threshold_____

Base excess: <2.0 ☐ 2.0-6.0 ☐ 6.1-10.0 ☐ 10.1-12.0 ☐ 12.1-16 ☐ >16 ☐

Median _____ Mean_____ Unreported ☐

Reference Standard / Outcome

25) Measured blind from diagnostic test: Yes ☐ No ☐ Unclear ☐

Reference standard used:_____

_____

Threshold_____

Dataset used to establish threshold_____

_____

Appendix 4. Data extraction form for review of cord pH and base excess and adverse outcome

26)**Timing of measurement:** Age(yrs): Range_____
27) Schools:
Mainstream ☐ Special needs ☐ Mixed ☐ Unreported ☐ Other ☐

Results

Population:		**Reference Test: Threshold:**		
Index test, Measurement:		Positive	Negative	**Total**
	Positive	**TP**	**FP**	
Threshold:	Negative	**FN**	**TN**	
	Total			

Appendix 5. Data extraction form for review of birth weight standards and adverse outcome

Data Collection Sheet Birth weight standard

Date: _____

Reviewer ID:_____ **Paper No:**_____

Year Of Publication:_____ **Language:**_____

Region study performed:_____

Section A: Study Selection

Population:

Inclusion: Infants born at term (≥ 37 weeks) gestation

Yes ☐ (Include) No ☐ (study excluded)

Index test:

Birth weight, ponderal index or other growth measurements performed:

Yes ☐ (include) No ☐ (study excluded)

Reference Test / Outcome Measure:

A reference standard looking at childhood or adult morbidity is performed

Yes ☐ (include) No ☐ (study excluded)

2 x 2 Table Possible:

Yes ☐ (include) No ☐ (exclude and case series ≤ 10, if other data e.g. sens, spec can include)

Study Selected:

Yes ☐ (must answer yes to ALL include questions)

No ☐ Give reason if NO _____

1)Ref ID:		**4)Publication year:**	
2)Rev name:		**5)First Author:**	

Appendix 5. Data extraction form for review of birth weight standards and adverse outcome

3)Country:		6)Language:	

Section B: Data Retrieval for Birth weight standards review

Population

7) Healthcare Centre:

Secondary care ☐ Tertiary care ☐ Mixed ☐ Other ☐ Unreported ☐

8) Setting:

In-patient ☐ Unreported ☐ Other ☐_____

9) Number of participating centres: _____

10) Gestation at time of delivery:

34-37 weeks ☐ 37-40 weeks ☐ >40 weeks ☐ Unreported ☐ Other _____

10.i) Mean (range)_____ Unreported ☐

10.ii) Median (range) _____ Unreported ☐

10.iii)

11) Pregnancy:

Low Risk ☐ High Risk ☐ Unselected ☐ Unreported ☐

11.i) State high risk conditions: Unreported

12) Did all mothers have singleton pregnancies?:

Yes ☐ No ☐ Unreported ☐

13) Were mothers primigravid?:

Yes ☐ No ☐ Unreported ☐

14) What was the mode of delivery?

Appendix 5. Data extraction form for review of birth weight standards and adverse outcome

SVD ☐ Instrumental ☐ Emergency Caesarean section ☐

Elective Caesarean section ☐ Mixed ☐ Unreported ☐

15) Index test: < 3rd centile ☐ < 5th centile ☐ < 10th centile ☐ < 25th centile ☐

> 2SD ☐

<500g ☐ <1000g ☐ <1500g ☐ <2000g ☐ <2500g ☐

Other ☐ _____ Unclear ☐

Unreported ☐

Timing of measurement_____

Method of measurement_____

No. of operators/ experience_____

16) List other eligibility/ in-/exclusion criteria:

17) Study population: (describe age (mean +/- SD or median/range), ethnicity, social class, breastfeeding etc.

Appendix 5. Data extraction form for review of birth weight standards and adverse outcome

18) Start of patient inclusion (year) : ☐ Unreported ☐

19) End of patient inclusion (year) : ☐ Unreported ☐

20) Study Design:

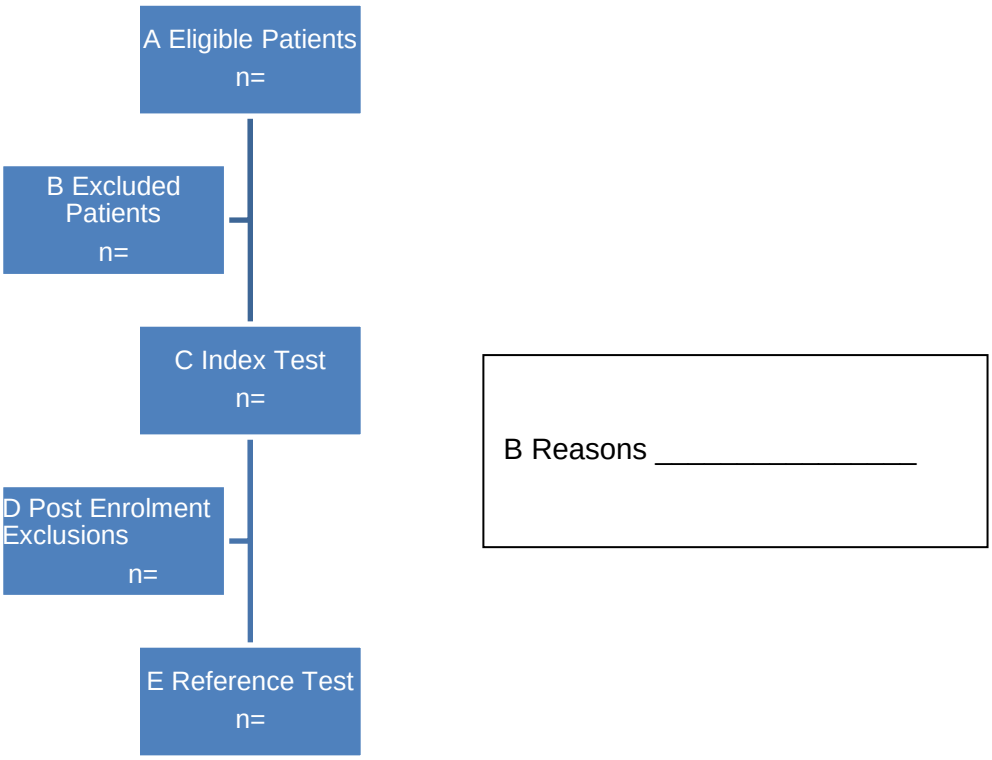
cohort ☐ case control ☐ RCT/CCT ☐ cross sectional ☐ before and after ☐

case series ☐ (no _____) other ☐

20.i) Data collection: prospective ☐ retrospective ☐ unreported ☐ other ☐

20.ii) Enrolment: consecutive ☐ arbitary (random) ☐ unreported ☐ other ☐

21) Numbers:



21) Completeness of Verification:

(= E / C x 100 = %)

> 90% ☐ 81-90% ☐ < 81% ☐

Index Test

22) Description of technique:

Adequate ☐ Inadequate ☐

15) Index test:

Birthweight < 3rd centile ☐ < 5th centile ☐ < 10th centile ☐ < 25th centile ☐

<2SD ☐

<500g ☐ <1000g ☐ <1500g ☐ <2000g ☐ <2500g ☐

Other ☐ _____ Unclear ☐

Unreported ☐

Timing of measurement_____

Method of measurement_____

No. of operators/ experience_____

Reference Standard / Outcome

25) Measured blind from diagnostic test: Yes ☐ No ☐ Unclear ☐

Reference standard used:_____

_____

Threshold_____

Dataset used to establish threshold_____

_____

26)Timing of measurement: Age(yrs): Range_____

Appendix 5. Data extraction form for review of birth weight standards and adverse outcome

Results

Population:		**Reference Test:** **Threshold:**		
Index test, Measurement: **Threshold:**		Positive	Negative	**Total**
	Positive	**TP**	**FP**	
	Negative	**FN**	**TN**	
	Total			

Other Information: (ie. other statistics, measures of uncertainty etc)

Appendix 6. Data collection sheet for review of Apgar score and adverse outcomes

Data Collection Sheet Apgar

Date: _____

Reviewer ID:_____ **Paper No:**_____

Year Of Publication:_____ **Language:**_____

Region study performed:_____

Section A: Study Selection

Population:

Inclusion: Infants with apgar score performed at birth

Yes ☐ (Include) No ☐ (study excluded)

Index test:

Apgar score performed:

Yes ☐ (include) No ☐ (study excluded)

Reference Test / Outcome Measure:

A reference standard looking at neonatal, childhood or adult morbidity/mortality is performed

Yes ☐ (include) No ☐ (study excluded)

2 x 2 Table Possible:

Yes ☐ (include) No ☐ (exclude and case series ≤ 10, if other data e.g. sens, spec can include)

Study Selected:

Yes ☐ (must answer yes to ALL include questions)

No ☐ Give reason if NO _____

1)Ref ID:		4)Publication year:	

Appendix 6. Data collection sheet for review of Apgar score and adverse outcomes

2)Rev name:		**5)First Author:**	
3)Country:		**6)Language:**	

Section B: Data Retrieval for Apgar Study

Population

7) Healthcare Centre:

Secondary care ☐ Tertiary care ☐ Mixed ☐ Other ☐ Unreported ☐

8) Setting:

In-patient ☐ Unreported ☐ Other ☐_____

9) Number of participating centres: _____

10) Gestation at time of delivery:

20-24 weeks ☐ 24-28 weeks ☐ 28-34 weeks ☐ 34-37 weeks ☐ 37-40 weeks ☐

>40 weeks ☐ Unreported ☐ Other _____

10.i) Mean (range)_____ Unreported ☐

10.ii) Median (range) _____ Unreported ☐

10.iii)

11) Pregnancy:

Low Risk ☐ High Risk ☐ Unselected ☐ Unreported ☐

11.i) State high risk conditions: Unreported

12) Did all mothers have singleton pregnancies?:

Yes ☐ No ☐ Unreported ☐

13) Were mothers primigravid?:

Yes ☐ No ☐ Unreported ☐

Appendix 6. Data collection sheet for review of Apgar score and adverse outcomes

14) What was the mode of delivery?

SVD ☐ Instrumental ☐ Emergency Caesarean section ☐

Elective Caesarean section ☐ Mixed ☐ Unreported ☐

15) Birth weight: < 3rd centile ☐ < 5th centile ☐ < 10th centile ☐ < 25th centile ☐

> 2SD ☐

<500g ☐ <1000g ☐ <1500g ☐ <2000g ☐ <2500g ☐ 2500-4000g ☐ >4000g ☐

Other ☐ _____ Unclear ☐

Unreported ☐

Timing of measurement_____

Method of measurement_____

No. of operators/ experience_____

16) List other eligibility/ in-/exclusion criteria:

17) Study population: (describe age (mean +/- SD or median/range), ethnicity, social class, breastfeeding etc.

Appendix 6. Data collection sheet for review of Apgar score and adverse outcomes

18) Start of patient inclusion (year) : ☐ Unreported ☐

19) End of patient inclusion (year) : ☐ Unreported ☐

20) Study Design:

cohort ☐ case control ☐ RCT/CCT ☐ cross sectional ☐ before and after ☐

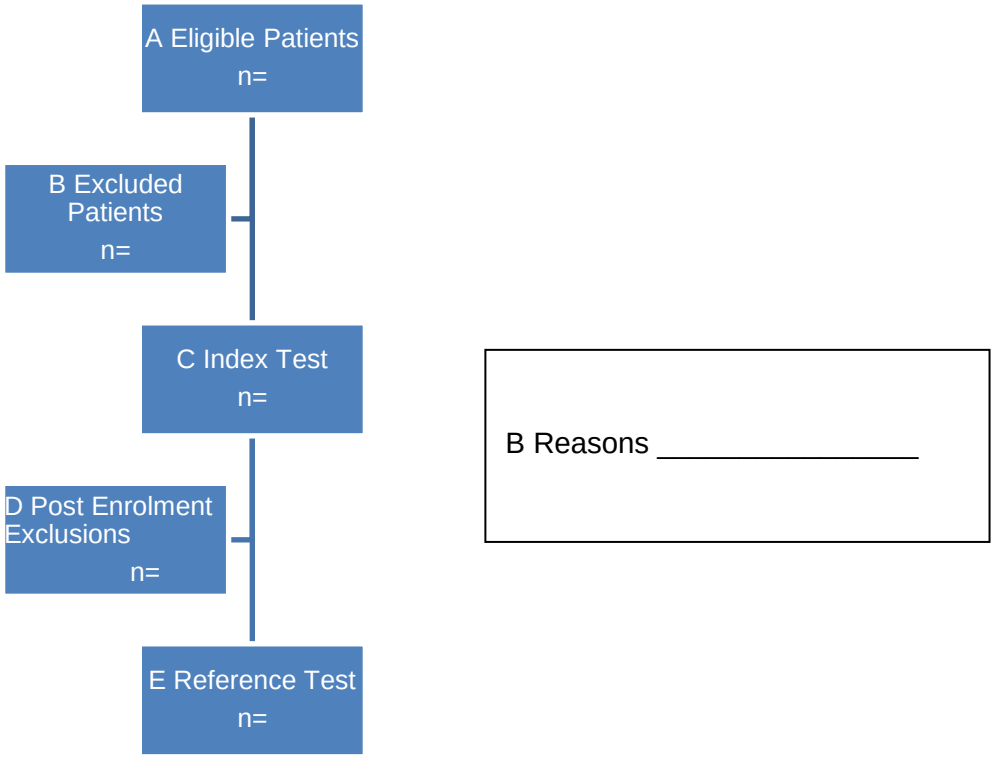
case series ☐ (no _____) other ☐

20.i) Data collection: prospective ☐ retrospective ☐ unreported ☐ other ☐

20.ii) Enrolment: consecutive ☐ arbitary (random) ☐ unreported ☐ other ☐

21) Numbers:

Appendix 6. Data collection sheet for review of Apgar score and adverse outcomes



21) Completeness of Verification:

(= E / C x 100 = %)

> 90% ☐ 81-90% ☐ < 81% ☐

Index Test

22) Description of technique:

Adequate ☐ Inadequate ☐

15) Index test:

Apgar score 1 min ☐ 5min ☐ 10 min ☐

Threshold: _____

Other ☐ _____ Unclear ☐

Unreported ☐

Timing of measurement_____

Appendix 6. Data collection sheet for review of Apgar score and adverse outcomes

Method of measurement_____

No. of operators/ experience_____

Reference Standard / Outcome

25) Measured blind from diagnostic test: Yes ☐ No ☐ Unclear ☐

Reference standard used:_____

_____

Threshold_____

Dataset used to establish threshold_____

_____

26)Timing of measurement: Age(yrs): Range_____

Results

Population:		**Reference Test: Threshold:**		
Index test, Measurement:		Positive	Negative	**Total**
	Positive	**TP**	**FP**	
Threshold:	Negative	**FN**	**TN**	
	Total			

Other Information: (ie. other statistics, measures of uncertainty etc)

Appendix 7. Medline search strategy for the association of umbilical cord pH and base excess with neonatal and long term outcomes

The following MESH headings (exp indicates selection expanded) and keywords (followed by 'mp') were used:

1. cord gases.mp.
2. cord pH.mp.
3. umbilical artery pH.mp.
4. umbilical cord blood.mp.
5. Blood Gas Analysis
6. exp Umbilical Cord
7. exp Hydrogen-Ion Concentration
8. Blood/ or blood.mp.
9. exp Infant, Newborn
10. Asphyxia Neonatorum
11. exp Brain Damage, Chronic
12. exp Hypoxia-Ischemia, Brain
13. exp Mental Disorders Diagnosed in Childhood
14. cerebral palsy.mp.
15. developmental delay.mp.
16. hypoxic ischaemic encephalopathy.mp.
17. handicap*.mp.
18. mental retard*.mp.
19. feeding difficulties.mp.
20. Infant Mortality
21. Child Mortality

The search terms were combined as follows:

22. 1 or 2 or 3 or 4 or 5
23. 6 and 7 and 8

Appendix 7. Medline search strategy for the association of umbilical cord pH and base excess with neonatal and long term outcomes

24. 22 or 23
25. 9 and 24
26. 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21
27. 25 and 26
28. Limit 27 to humans

Search results:

Medline 2983 citations

Embase 2701 citations

Web of Science 1179 citations

Cochrane library 5 citations

MEDION 0 citations

SIGLE 0 citations

Excluding duplicates Total = 5690

Appendix 8. Table of included studies for systematic review of the association of umbilical cord pH and base excess with neonatal and long term outcomes (grouped according to population risk)

Author	Study design	Population risk (total number)	Risk Factor	Index Test	Outcome Measure
Baenziger et al 1999 Switzerland	Prospective cohort	High (10)	Total number included=10 All ventilated neonates. Risk factors for HIE inc meconium liquor, CTG abnormalities, low Apgar or pH. > 34 weeks gestation	Arterial cord pH < 7.00	Death[a] HIE (Sarnat gd >1)[a] Neurological Optimality Score (age 1 yr) Griffiths Developmental Scale (age 1 yr)
Beeby et al 1994 Australia	Prospective cohort	High (623)	Gestation <32 weeks. Congenital anomalies excluded	Arterial cord pH < 7.1	Death[a] Cerebral Palsy (diagnostic criteria not specified) (age 1 yr) Intra-ventricular haemorrhage Gd 3 or 4 on cranial USS or at autopsy[a]
Blackwell et al 2001 USA	Retrospective cohort	High (48)	All required ventilation > 48 hrs for meconium aspiration syndrome Gestation >37 weeks	Arterial cord pH < 7.2	Seizures[a] (diagnostic criteria not described)
Bresadola et al 1995 Italy	Cohort, unclear if prospective/ retrospective	High (452)	Gestation > 24 weeks < 37 weeks Birth weight >700g Congenital anomalies excluded	Arterial cord pH < 7.2	Death[a]
Casey et al 2001 USA	Retrospective cohort	High (1691)	Neonates who developed respiratory symptoms postnatally Gestation >37 weeks	Arterial cord pH < 7.2	Death[a] Seizures(1st 24 hours)[a] Respiratory Distress (requiring ventilation)[a] Meconium aspiration syndrome (radiographically diagnosed)[a]
Engle 1999 USA	Retrospective cohort	High (73)	Neonates admitted to Neonatal unit directly from Delivery suite Gestation > 37 weeks	Arterial cord pH < 7.00	HIE including seizures[a] (threshold unreported)

Appendix 8. Table of included studies for systematic review of the association of umbilical cord pH and base excess with neonatal and long term outcomes (grouped according to population risk)

Ertan 2001 Germany	Case control study	High (120)	Gestation 24-34 weeks	Arterial cord pH <7.10 Arterial base excess (BE) > 16mmol/l	Intra-cranial haemorrhage (all grades- Papile) on cranial USS[a]
Gaudier et al 1994 USA	Retrospective cohort	High (216)	Birth weight 500-1000g Gestation 23-34 weeks Congenital anomalies excluded	Arterial cord pH < 7.05	Weschler intelligence scale for children (IQ < 70) Cerebral palsy (chronic non progressive motor disability characterised by abnormal posture and movements)(follow up to age 7 years) Motor or mental deficit severe enough to interfere with normal function including one or more of mental retardation, cerebral palsy, deaf, blind, hydrocephaly
Gea 2007 Brazil	Prospective cohort	High (25)	Birth weight < 2000g Gestation < 37 weeks Excluded congenital anomalies, maternal diabetes and rhesus incompatibility	Venous pH < 7.20 Venous BE > 10mmol/l	Periventricular or intra-ventricular haemorrhage on cranial USS[a] Ventilation > 24 hours[a] Necrotising enterocolitis (Grade 2 Bell's Classification)[a]
Gonzalez de Dios et al 2000 Spain	Prospective cohort	High (180)	At least one risk factor for asphyxia e.g. Apgar <6, cord pH < 7.00, antenatal risk factors. Exc if congenital anomaly, sepsis, metabolic disorder, PN depression Gestation > 37 weeks	Arterial cord pH ≤ 7.00	Serum creatinine >1.2mg/dl[a,c] HIE(Levine's criteria all grades)[a] Abnormal neurological status (Amiel-Tison criteria) (age 2 years)

Appendix 8. Table of included studies for systematic review of the association of umbilical cord pH and base excess with neonatal and long term outcomes (grouped according to population risk)

Graham et al 2004 USA	Case control study	High (220)	Gestation 23-34 weeks	Arterial cord pH <7.00	Periventricular leukomalacia or ventricular dilatation on cranial USS[a,b]
Haddad et al 2000 USA	Retrospective cohort	High (28)	Neonates with Apgar 0 at 1 and 5 mins resuscitated and tranferred to Neonatal unit, excluding congenital malformations, chromosone abnormalities, Birth before arrival Gestation 25.5-42.1 wks	Arterial cord pH <7.00	Death[a] HIE[a] (diagnostic criteria not described)
Hernandez et al 1993 USA	Retrospective cohort	High (82)	Clinical and radiological evidence of meconium aspiration syndrome. Excluded if congenital anomaly, cytomegalic inclusion disease, delivery outside hospital	Arterial cord pH <7.00 Arterial cord pH <7.10 Arterial cord pH <7.20	Ventilation required[a] Ventilation ≥ 3 days[a]
Hibbard et al 1991 USA	Retrospective cohort	High (171)	Birth weight 500-1500g	Arterial cord pH ≤ 7.05 Arterial cord pH ≤ 7.15	Death[a] Intraventricular haemorrhage (Gd 1-4) on Cranial USS[a] Abnormal neurological status defined as seizures, cortical atrophy, need for shunt placement[a] Hyaline membrane disease (CXR)[a] Bronchopulmonary dysplasia (CXR)[a] Necrotising enterocolitis[a]
Holmes et al 2001 Canada	Case control study	High (76)	BW 750-2500g. Gestation 25-35 weeks Excluded if CS before labour, congenital anomalies, uninterpretable FHR trace	Arterial cord pH <7.10	Death[a] Intraventricular haemorrhage (Grade 3 or 4 Papile)[a] Periventricular leukomalacia[a]

Appendix 8. Table of included studies for systematic review of the association of umbilical cord pH and base excess with neonatal and long term outcomes (grouped according to population risk)

Kato 1997 Japan	Retrospective cohort	High (195)	Birth weight <1500g. Congenital anomalies excluded	Arterial cord pH < 7.20	Death[a] Cerebral palsy or mental retardation (threshold unreported) (at least 12 months old)
Loh et al 1998 Singapore	Prospective cohort	High (69)	Included if one or more of the following risk factors: CTG abnormality, scalp pH< 7.25, thick mec/no liquor, cord prolapse/ bradycardia, antepartum haemorrhage, EFW<1.5kg, <34 weeks, breech, poorly controlled IDDM, PET, suspected fetal anomalies, transverse/oblique lie at CS, multiple preg	Arterial cord pH <7.00	HIE (diagnostic criteria unreported)
Luthy et al 1987 USA	Prospective cohort	High (199)	Gestation 26 – 32 weeks Excluded multiple pregnancy, non-cephalic presentation, malformations, delivered before labour, antenatal haemorrhage	Arterial cord pH ≤ 7.20	Death[a] Cerebral palsy (diagnostic criteria unreported) (18 months of age) IVH on cranial USS (Grade 3 / 4 Papile)[a]
Murphy et al 1995 UK	Case control study	High (152)	Gestation 23-32 weeks. Excluded multiple pregnancy	Arterial cord pH ≤ 7.10	Cerebral palsy (permanent impairment of voluntary movement or posture)(age unreported)
Salafia et al 1995 USA	Retrospective cohort	High (406)	Gestation < 32 weeks. Excluded if congenital anomalies, multiple pregnancy, maternal diabetes mellitus or chronic hypertension, fetal hydrops, placenta praevia, intra-uterine growth restriction	Arterial cord pH <7.10 Venous cord pH <7.10	Germinal matrix or intraventricular haemorrhage diagnosed on cranial USS(Grade 1-4)[a]

Appendix 8. Table of included studies for systematic review of the association of umbilical cord pH and base excess with neonatal and long term outcomes (grouped according to population risk)

Socol 1994 USA	Retrospective cohort	High (28)	APGAR score ≤3 at 5 minutes. Excluded birthweight < 2000g and gestation < 34 weeks	Arterial cord pH <7.00 Arterial base excess >12mmol/l	Cerebral palsy / motor deficit (diagnostic criteria unreported) (age 1-7 years) Seizures[a] Renal impairment (Serum creatinine >1.5mg/dl or oliguria)[a] Cognitive impairment on Welscher Scale (cut off <70) (age 1-7 years)
Spinillo et al 1995 Italy	Case control study	High (159)	Birth weight < 2500g	Arterial cord pH <7.20	Bayley Index abnormal (71-84). (Age 2 years)
Tejani and Verma 1989 USA	Cohort, unclear if prospective/ retrospective	High (392)	Birth weight ≤2000g. Excluded major congenital anomalies	Arterial cord pH ≤ 7.10	Death[a] Intraventricular haemorrhage on cranial USS (Gd 1-4 Papile)[a] RDS (radiological evidence of reticulogranular pattern and air bronchograms)[a]
Yudkin et al 1994 UK	Retrospective cohort	High (122)	APGAR score ≤3 at 1 minute Gestation >37 weeks Excluded multiple pregnancies and death related to congenital anomaly or rhesus disease	Arterial cord pH < 7.15	Death[a] Any impairment (age 5 years) Serious impairment (global delay, lateralising signs and severe deficit in a specific area) (age 5 years)
Yoon et al 1996 Korea	1996	High (153)	Gestation 25 – 36 weeks Excluding major congenital malformations or death before examination	Arterial cord pH < 7.15	Periventricular leukomalacia on cranial USS (cystic lesions in PV white matter or bulging ventricle adjacent to white matter or persistent increase in echogenicity) or at autopsy[a]

Appendix 8. Table of included studies for systematic review of the association of umbilical cord pH and base excess with neonatal and long term outcomes (grouped according to population risk)

Dennis et al 1989 UK	Retrospective cohort	Unselected (189)	Gestation > 37 weeks Singletons surviving to age 4.5 years	Arterial cord pH ≤ 7.10 Arterial Base Excess > 12 mmol/l	Griffiths Developmental scales age 4.5 years(<10th percentile): Locomotor Personal/social Hearing/speech Performance Overall impairment
Dijxhoorn et al 1986 Netherlands	Retrospective cohort	Low (803)	Gestation > 37 weeks Excluding caesarean delivery or breech presentation	Arterial cord pH ≤ 7.10 Arterial cord pH ≤ 7.20	Neurological status (abnormal if one of hyper/hypokinesia, hyper/hypotonia, hemisyndrome, apathy syndrome, hyperexcitability syndrome)[a]
D'Souza et al 1983 UK	Cohort unclear if prospective/ retrospective	Low (453)	Normal pregnancy, vaginal delivery 39-42 weeks	Venous pH < 7.27	Neurological status (abnormal if one of hypotonia, lethargy, feeding difficulties, jittery)[a]
Ghosh et al 2003 India	Prospective cohort	Low (75)	Gestation >37 weeks Singletons excluded if Rhesus isoimmunisation, maternal anaemia or diabetes mellitus	Arterial pH ≤ 7.15	HIE(threshold unreported)[a] Death[a]
Gilstrap et al 1989 USA	Retrospective cohort	Low (2738)	Gestation > 37 weeks, cephalic presentation, birth weight > 2500g Excluding congenital anomalies	Arterial cord pH < 7.00	Hypotonia > 24-48 hours[a] Respiratory disease requiring oxygen[a] Seizures[a]
Graham 2002 USA	Case control study	Unselected (34)	Excluding congenital anomalies	Arterial cord pH <7.00	Seizures[a]

Appendix 8. Table of included studies for systematic review of the association of umbilical cord pH and base excess with neonatal and long term outcomes (grouped according to population risk)

Heller et al 2003 Germany	Retrospective cohort	Unselected (464345)	Excluding congenital anomalies	Arterial cord pH ≤ 7.00 Arterial cord pH ≤ 7.10 Arterial cord pH ≤ 7.20	Death[a]
Hogan 2007 Sweden	Case control study	Unselected (313)	Gestation >37 weeks	Arterial cord pH ≤ 7.15	Composite reference standard HIE(Sarnat all grades) or hypoxic death[a]
Huisjes and Aarnoudse 1979 Netherlands	Cohort unclear if prospective/ retrospective	Unreported (838)	Population characteristics unreported	Arterial cord pH ≤ 7.09	Neonatal neurological status abnormal (Prechtl)[a]
Ingemarrson et al 1997 Sweden	Case control study	Unreported (308)	Population characteristics unreported	Arterial cord pH <7.05 Arterial cord pH <7.00	Death[a] Cerebral palsy (age 4 years) HIE (Sarnat Grade 1-3)[a] Attention deficit (age 4 yrs) Speech difficulty (age 4 yrs) Motor delay (age 4 yrs)
Jurgens- van der Zee et al 1979 Netherlands	Prospective cohort	Unselected (1343)	Neonates. Excluded if neonatal death or parents refused examination	Venous pH < 7.20	Neuro exam abnormal (Prechtl): 1 or more of increased/decreased excitability, seizures, apathy, coma. Abnormal motility or tone. Peripheral/central nervous system lesions[a]
Larma et al 2007 USA	Case control study	Unselected (214)	Gestation ≥ 24 weeks	Arterial cord pH < 7.00	Seizures[a] Periventricular leukomalacia[a] Intra-ventricular haemorrhage[a] Respiratory dysfunction[a] Renal dysfunction[a] (thresholds unreported)
Litschgi et al 1974 Germany	Prospective cohort	Unreported (1000)	Population characteristics unreported	Arterial cord pH < 7.09	Neurological examination abnormal (24 hours after birth)[a]

Appendix 8. Table of included studies for systematic review of the association of umbilical cord pH and base excess with neonatal and long term outcomes (grouped according to population risk)

Low 1997 Canada	Case control study	Unselected (174)	Gestation ≥ 37 weeks	Arterial base excess > 12mmol/l	Composite outcome score[a] Neurological abnormality (lethargy/ abnormal tone/coma/seizures)[a] Respiratory dysfunction (CPAP or ventilation required)[a] Cardiovascular dysfunction (hypo/hypertension/abnormal ECG or echocardiogram)[a] Renal dysfunction (Serum creatinine >100umol/l /anuria / oliguria <1ml/kg/hr)[a]
Perlman and Risser 1996 USA	Prospective cohort	Unselected (96)	Gestation ≥ 37 weeks	Arterial cord pH ≤ 7.00	Seizures[a]
Sakuraba and Saling 1989 Germany	Prospective cohort	Unreported (178)	Population characteristics unreported	Arterial cord pH ≤ 7.19 Arterial cord pH ≤ 7.24	Intracranial haemorrhage diagnosed on cranial USS(threshold unreported)[a,b]
Schneider and Tanner 1985 Germany	Retrospective cohort	Unselected (29)	Twins only included	Arterial cord pH <7.20	Binet Simon Kramer (intelligence) Language test Emotional intelligence test Raven Intelligence test(non-spoken) Neurological status (All tests performed age 5-7 years)
Silva et al 2008 USA	Case control study	Unselected (174)	Gestation ≥ 34 weeks. Congenital malformations and chromosome anomalies excluded	Arterial cord pH < 7.00 Arterial cord pH < 7.10	Hypotonia at birth necessitating NNU admission[a] HIE (threshold unreported)[a] Seizures[a]

Appendix 8. Table of included studies for systematic review of the association of umbilical cord pH and base excess with neonatal and long term outcomes (grouped according to population risk)

Svirko et al 2008 UK	Retrospective cohort	Unselected (87)	Gestation≥ 36 weeks. Included if had one of reference tests performed and cord pH available. Excluded if delivered by pre-labour elective CS	Arterial cord pH < 7.10	Weschler Objective Reading Dimensions (WORD) (age 6-8 yrs) Test for Comprehension of Grammar (TROG) (age 5-7 yrs) Naglieri non-verbal ability (NNAT) (age 6-8 yrs) (cut off <100 age standardised score used for all tests)
Thoulon et al 1972 France	Cohort unclear if prospective/ retrospective	Unreported (487)	Population characteristics unreported	Venous cord pH < 7.20	Death[a] Abnormal neurological examination (threshold unreported)(age 18months)
Valentin et al 1993 Sweden	Cohort unclear if prospective/ retrospective	Unselected (178)	Anomalies not excluded	Arterial cord pH ≤ 7.00 Arterial cord pH ≤ 7.10 Venous cord pH ≤ 7.10 Venous cord pH ≤ 7.20	Composite measure of neonatal sequelae including severe symptoms requiring treatment e.g ventilation, IV fluids or death or survival with sequelae[a]
Van den Berg 1996 Netherlands	Retrospective cohort	Unselected (168)	Excluded chromosomal or major congenital anomalies or intrauterine infection	Arterial cord pH < 7.00 (compared to group of neonates with cord pH > 7.24)	Seizures[a] Intracranial haemorrhage on Cranial USS[a,b] Periventricular leukomalacia on cranial USS[a] Renal impairment (serum creatinine > 90[a] Abnormal liver function (AST >33U/L, ALT >25U/L)[a,e] Necrotising enterocolitis (criteria unreported)[a]

Appendix 8. Table of included studies for systematic review of the association of umbilical cord pH and base excess with neonatal and long term outcomes (grouped according to population risk)

Vintzileos et al Greece	Prospective cohort	Unselected (678)	Gestation ≥26 weeks Excluding known congenital or chromosomal anomalies	Arterial cord pH <7.10	Death[a]
Wildshut et al 2005 Netherlands	Prospective cohort	Low (44)	Gestation 37- 42 weeks Neonates included if growth between 2.3-97.7th percentiles, vertex position, stay in hosp at least 3 days after birth. Excluded if HIE if caused by meconium aspiration, RDS, infection, born after complicated pregnancy, congenital malformations, maternal medication, alcohol or drug use, metabolic disorders.	Arterial cord pH < 7.10	Movement ABC test (score < 16 or unable to perform due to movement disorder) (age 4 years)
Winkler 1991 USA	Case control study	Unselected (713)	Gestation ≥ 37 weeks	Arterial cord pH < 7.20	Composite reference standard: seizures or neonatal death[a]
Wu et al 1998 Chinas	Case control study	Unselected (194)	Gestation ≥ 37 weeks Singletons only	Arterial cord pH < 7.19	Motor delay Speech delay Difficulty concentrating (all assessed age 4 years)

a=outcomes within the neonatal period

USS= ultrasound scan, Mg/dl= milligrams per decilitre, CXR= chest x-ray, U/L = units per litre, RDS= respiratory distress syndrome, HIE= hypoxic ischaemic encephalopathy, CS= Caesarean section, CPAP= continuous positive airway pressure, NNU= neonatal unit, ECG= electrocardiogram, AST= aspartate transaminase, ALT= alanine aminotransferase, Gd= Grade, IQ= intelligence quotient

Appendix 9. Reference list of included studies in systematic review of umbilical cord pH and neonatal and long term outcomes

Baenziger O, Moenkhoff M, Morales CG, Waldvogel K, Wolf M, Bucher H et al. Impaired chemical coupling of cerebral blood flow is compatible with intact neurological outcome in neonates with perinatal risk factors. *Biology of the Neonate* 1999; 75(1):9-17.

Beeby PJ, Elliott EJ, Henderson-Smart DJ, Rieger ID. Predictive value of umbilical artery pH in preterm infants. *Archives of Disease in Childhood* 1994; 71(2):F93-F96.

Blackwell SC, Moldenhauer J, Hassan SS, Redman ME, Refuerzo JS, Berry SM et al. Meconium aspiration syndrome in term neonates with normal acid-base status at delivery: is it different? *American Journal of Obstetrics & Gynecology* 2001; 184(7):1422-1425.

Bresadola M, Lo Mastro M, Arena V, Bellaveglia L, Scarpellini F. Preterm labour and neonatal parameters. *Clinical & Experimental Obstetrics & Gynecology* 1995; 3: 235-239.

Casey BM, Goldaber KG, McIntire DD. Outcomes amongst term infants when two-hour postnatal pH is compared with pH at delivery. *American Journal of Obstetrics & Gynecology* 2001; 184: 447-450.

D'Souza SW, Black P, Cadman J, Richards B. Umbilical venous blood pH: a useful aid in the diagnosis of asphyxia at birth. *Archives of Disease in Childhood* 1983; 58(1):15-19.

Dennis J, Johnson A, Mutch L, Yudkin P, Johnson P. Acid-base status at birth and neurodevelopmental outcome at four and one-half years. *American Journal of Obstetrics & Gynecology* 1989; 161(1):213-220.

Dijxhoorn MJ, Visser GH, Fidler V, Touwen BC, Huisjes HJ. Apgar score meconium and acidaemia at birth in relation to neonatal neurological morbidity in term infants. *British Journal of Obstetrics and Gynaecology* 1986; 93: 217-22.

Engle D, Laptook AR, Perlman J. Acute changes in arterial carbon dioxide tension and acid-base status and early neurologic characteristics in term infants following perinatal asphyxia. *Resuscitation* 1999; 42(1):11-17.

Ertan AK, Tanriverdi HA, Meier M, Schmidt W, Ertan AK, Tanriverdi HA et al. Perinatal risk factors for neonatal intracerebral hemorrhage in preterm infants. *European Journal of Obstetrics, Gynecology, & Reproductive Biology* 2006; 127(1):29-34.

Gaudier FL, Goldenberg RL, Nelson KG, Peralta-Carcelen M, Johnson SE, DuBard MB et al. Acid-base status at birth and subsequent neurosensory impairment in surviving 500 to 1000 gm infants. *American Journal of Obstetrics & Gynecology* 1994; 170(1 Pt 1):48-53.

Appendix 9. Reference list of included studies in systematic review of umbilical cord pH and neonatal and long term outcomes

Gea YA. Clinical value of lactate measurement and nucleated red blood cell counts in the placental segment of the umbilical vein of premature newborns for diagnosis of hypoxia-ischemia. *Jornal de Pediatria* 2007; 83(2):186-90.

Ghosh BM. Prediction of perinatal asphyxia with nucleated red blood cells in cord blood of newborns. *International Journal of Gynecology and Obstetrics* 2003; 81(3): 267-271.

Gilstrap LC, III, Leveno KJ, Burris J, Williams ML, Little BB. Diagnosis of birth asphyxia on the basis of fetal pH, Apgar score, and newborn cerebral dysfunction. *American Journal of Obstetrics & Gynecology* 1989; 161(3):825-830.

Gonzalez de DJ, Moya M, Carratala F. Neurological evolution of asphyctic full term newborns with severe umbilical acidosis (pHUA <7.00). *Revista de Neurologia* 2000; 31(2):107-113.

Graham EMH. Holcroft CJ, Blakemore KJ. Evidence of intrapartum hypoxia ischemia is not present in the majority of cases of neonatal seizures. *Journal of Maternal-Fetal and Neonatal Medicine* 2002; 12(2):123-6.

Graham EMH, Holcroft CJ, Karishma KR, Donahue PK, Allen MC. Neonatal cerebral white matter injury in preterm infants is associated with culture positive infections and only rarely with metabolic acidosis. *American Journal of Obstetrics and Gynecology* 2004; 191(4):1305-10.

Haddad B, Mercer BM, Livingston JC, Talati A, Sibai BM. Outcome after successful resuscitation of babies born with apgar scores of 0 at both 1 and 5 minutes. *American Journal of Obstetrics & Gynecology* 2000; 182(5):1210-1214.

Heller G, Schnell RR, Misselwitz B, Schmidt S. Umbilical blood pH, Apgar scores, and early neonatal mortality. *Zeitschrift fur Geburtshilfe und Neonatologie* 2003; 207(3):84-89.

Hernandez C, Little BB, Dax JS, Gilstrap LC, III, Rosenfeld CR. Prediction of the severity of meconium aspiration syndrome. *American Journal of Obstetrics & Gynecology* 1993; 169(1):61-70.

Hibbard JU, Hibbard MC, Whalen MP. Umbilical cord blood gases and mortality and morbidity in the very low birth weight infant. *Obstetrics & Gynecology* 1991; 78(5 Pt 1):768-773.

Hogan LI. How often is a low 5-min Apgar score in term newborns due to asphyxia? *European Journal of Obstetrics Gynecology and Reproductive Biology* 2007; 130(2): 169-175.

Holmes P, Oppenheimer LW, Gravelle A, Walker M, Blayney M. The effect of variable heart rate decelerations on intraventricular hemorrhage and other perinatal outcomes in preterm infants. *Journal of Maternal-Fetal Medicine* 2001; 10(4):264-268.

Appendix 9. Reference list of included studies in systematic review of umbilical cord pH and neonatal and long term outcomes

Huisjes HJ, Aarnoudse JG. Arterial or venous umbilical pH as a measure of neonatal morbidity? *Early Human Development* 1979; 3(2):155-161.

Ingemarsson I, Herbst A, Thorngren-Jerneck K. Long term outcome after umbilical artery acidaemia at term birth: influence of gender and duration of fetal heart rate abnormalities. *British Journal of Obstetrics & Gynaecology* 1997; 104(10):1123-1127.

Jurgens-van der Zee AD, Bierman-van Eendenburg MEC, Fidler V, Olinger AA, Visch JH, Towen BCL et al. Preterm birth, growth retardation and acidemia in relation to neurological abnormality of the newborn. *Early Human Development* 1979; 32: 141-154.

Kato EHY, Yamada H, Matsumoto Y, Hattori S, Makinoda S, Fujimoto S. Relation between perinatal factors and outcome of very low birth weight infants. *Journal of Perinatal Medicine* 1996; 24(6):677-686.

Larma JD, Silva AM, Holcroft CJ, Thompson RE, Donohue PK, Graham EM et al. Intrapartum electronic fetal heart rate monitoring and the identification of metabolic acidosis and hypoxic-ischemic encephalopathy. *American Journal of Obstetrics & Gynecology* 2007; 197(3):301-308.

Litschgi M, Benz JJ, Glatthaar E. Actual and prognostic value of arterial cord pH for the newborn infant. *Zeitschrift fur Geburtshilfe und Perinatologie* 1974; 178:23-29.

Loh SF, Woodworth A, Yeo GS, Loh SF, Woodworth A, Yeo GS. Umbilical cord blood gas analysis at delivery. *Singapore Medical Journal* 1998; 39(4):151-155.

Low JA. Threshold of metabolic acidosis associated with newborn complications. *American Journal of Obstetrics & Gynecology* 1997; 177:1391-1394.

Luthy DA, Shy KK, Strickland D, Wilson J, Bennett FC, Brown ZA et al. Status of infants at birth and risk for adverse neonatal events and long-term sequelae: a study in low birth weight infants. *American Journal of Obstetrics & Gynecology* 1987; 157(3):676-679.

Murphy DJ, Sellers S, MacKenzie IZ, Yudkin P, Johnson A. Case-control study of antenatal and intrapartum risk factors for cerebral palsy in very preterm singleton babies. *Lancet* 1995; 346:1449-1454.

Perlman JM, Risser R. Can asphyxiated infants at risk for neonatal seizures be rapidly identified by current high-risk markers? *Pediatrics* 1996; 97(4):456-462.

Sakuraba M, Saling E. Umbilical cord blood coagulability, acidosis and intracranial hemorrhage. *Journal of Perinatal Medicine* 1989; 17(2):99-106.

Salafia CM, Minior VK, Rosenkrantz TS, Pezzullo JC, Popek EJ, Cusick W et al. Maternal, placental, and neonatal associations with early germinal

matrix/intraventricular hemorrhage in infants born before 32 weeks' gestation. *American Journal of Perinatology* 1995; 12(6):429-436.

Schneider R, Tanner R, Schneider R, Tanner R. Perinatal umbilical artery pH and cerebral function disorders in twins starting school. *Zeitschrift fur Kinder- und Jugendpsychiatrie* 1985; 13(1):24-30.

Silva AM, Cootauco AC, ina-Mumuney A, Donohue PK, Graham EM. The association of hypotonia and depression in the term and near-term neonate with metabolic acidemia. *Journal of Perinatal Medicine* 2008; 36(2):151-156.

Socol ML. Depressed Apgar score, acid-base balance and neurologic outcome. *American Journal of Obstetrics & Gynecology* 1994; 170: 991-999.

Spinillo A, Fazzi E, Orcesi S. Perinatal factors and 2 year minor neurodevelopmental impairment in low birth weight infants. *Biology of the Neonate* 1995; 67(1):39-46.

Svirko E, Mellanby J, Impey L. The association between cord pH at birth and intellectual function in childhood. *Early Human Development* 2008; 84(1):37-41.

Tejani N, Verma U. Correlation of Apgar scores and umbilical artery acid-base status to mortality and morbidity in the low birth weight neonate. *Obstetrics & Gynecology* 1989; 73:597-600.

Thoulon JM, Varnier C, Faure M. Prognostic value of umbilical blood pH measurement in newborn infants at birth. *Lyon Medical* 1972; 227(8):699-702.

Valentin L, Ekman G, Isberg PE, Polberger S, Marsal K, Valentin L et al. Clinical evaluation of the fetus and neonate. Relation between intra-partum cardiotocography, Apgar score, cord blood acid-base status and neonatal morbidity. *Archives of Gynecology & Obstetrics* 1993; 253(2):103-115.

Van den Berg PPN. Neonatal complications in newborns with an umbilical artery pH <7.00. *American Journal of Obstetrics and Gynecoogyl* 1996; 175(5): 1152-7.

Vintzileos AM, Antsaklis A, Varvarigos I, Papas C, Sofatzis I, Montgomery JT. A randomized trial of intrapartum electronic fetal heart rate monitoring versus intermittent auscultation. *Obstetrics & Gynecology* 1993; 81:899-907.

Wildschut J, Feron FJ, Hendriksen JG, van HM, Gavilanes-Jiminez DW, Hadders-Algra M et al. Acid-base status at birth, spontaneous motor behaviour at term and 3 months and neurodevelopmental outcome at age 4 years in full-term infants. *Early Human Development* 2005; 81(6):535-544.

Winkler CL. Neonatal complications at term as related to the degree of umbilical artery acidaemia. *American Journal of Obstetrics & Gynecology* 1991;164: 637-641.

Appendix 9. Reference list of included studies in systematic review of umbilical cord pH and neonatal and long term outcomes

Wu L, Thorngren-Jerneck K, Ingemasson I. Different types of acidemia at birth, fetal heart rate patterns and infants outcome at four years of age. *Chung-Hua Fu Chan Ko Tsa Chih [Chinese Journal of Obstetrics & Gynecology]* 1998; 33(8):462-465.

Yoon BH, Romero R, Yang SH, Jun JK, Kim IO, Choi JH et al. Interleukin-6 concentrations in umbilical cord plasma are elevated in neonates with white matter lesions associated with periventricular leukomalacia. *American Journal of Obstetrics & Gynecology* 1996; 174(5):1433-1440.

Yudkin P, Johnson A, Clover LM, Murphy KW. Clustering of perinatal markers of birth asphyxia and outcome age 5 years. *BJOG: An International Journal of Obstetrics & Gynaecology* 1994; 101:774-781.

Appendix 10. Medline search strategy for systematic review of birth weight standards

1. Birth Weight/
2. birth weight.mp.
3. birth-weight.mp.
4. birthweight.mp.
5. Infant, Very Low Birth Weight/ or Infant, Low Birth Weight/
6. lbw.mp.
7. small-for-gestational-age.mp.
8. small for gestational age.mp.
9. small for date*.mp.
10. small for gestation*.mp.
11. sga.mp.
12. Infant, Small for Gestational Age/
13. intrauterine growth restrict*.mp.
14. intrauterine growth retard*.mp.
15. IUGR.mp.
16. Fetal Growth Retardation/
17. ponderal index.mp.
18. abdominal circumference.mp.
19. head circumference.mp.
20. exp body constitution/ or "body weights and measures"/
21. neonat*.mp.
22. infant*.mp.
23. newborn.mp.
24. 22 or 21 or 23
25. 11 or 7 or 2 or 1 or 16 or 13 or 6 or 3 or 9 or 12 or 14 or 15 or 8 or 4 or 10 or 5
26. 18 or 19 or 17 or 20
27. 24 and 26
28. 27 or 25
29. Diabetes Mellitus/
30. diabetes mellitus.mp.
31. Hypertension/ or hypertension.mp.
32. cardiovascular disease.mp. or Cardiovascular Diseases/
33. Metabolic Syndrome X/ or metabolic syndrome.mp.
34. cerebral palsy.mp. or Cerebral Palsy/
35. Developmental Disabilities/

Appendix 10. Medline search strategy for systematic review of birth weight standards

36. developmental delay.mp.
37. learning difficulties.mp.
38. Learning Disorders/
39. neonatal morbidity.mp.
40. Infant Mortality/
41. neonatal mortality.mp.
42. infant mortality.mp.
43. 35 or 33 or 32 or 39 or 40 or 36 or 41 or 42 or 38 or 34 or 30 or 37 or 29 or 31
44. 28 and 43
45. limit 44 to humans

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Author, Year and Country	Study design	Population (total number)	Method of defining gestational age	Birth weight parameter	Outcome Measure
Algert et al 2009 Australia	Retrospective cohort	(475455) Singleton All live births in state within study period Year of birth2000-2005 Excluded those who moved out of area, no others specified Ethnicity unreported	Unreported	Birth weight <2.5 kg	Type 1 Diabetes Mellitus (age 3-6 years) (ICD-10 classification coded from hospital records)
Als et al 1976 USA	Prospective cohort	(20) Unclear if singleton/ mixed Infants with IUGR and same number of control infants Year of birth No congenital anomalies included Ethnicity: all Caucasian	LMP and Dubowitz method	Ponderal Index <2.25	1. Brazelton motor score (10 days old), below average 3
Amigo et al 2010 Brazil/Chile	Retrospective cohort	(2793) Singleton Two cohorts 1. Stratified sample (3 social levels) of adults who were singletons born between June 1978 and May 1979. 2. Adult subjects aged 22-28 years randomly selected from a sampling frame of newborns registered jan 1974- dec 1978 in small agricultural area No exclusions specified Ethnicity unreported	Unreported	Absolute birth weight <2.5 kg	2. BMI ≥ 30 aged 20-28years 3. Fasting total cholesterol ≥ 5mmol/L aged 20-28 years

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Ananth and Vintzelios 2009 USA	Retrospective cohort	(18169349) Singleton General population in USA during study period Year of birth 1995-2004 Congenital anomaly excluded Ethnicity 77.4% white, 17.5% black,5.2% other	Clinical estimate (combination of USS, obstetrical and newborn examination)	Population chart <10th centile (internal chart based on this population)	Neonatal death (28 days, from national dataset)
Andersson et al 2000 Sweden	Retrospective cohort	(438) Singleton General, all women born in years 1918,1922 and 1930 residing in city invited to participate No exclusions specified Ethnicity unreported	LMP	Birth weight <3.1kg	Hypertension at age 50 and age 60 years (under treatment and/ or BP ≥ 160mmHg systolic and/ or diastolic >95mmHg)
Arora, Paul and Singh 1987 India	Prospective cohort	(200) Unclear if singleton/ mixed Consecutive hospital born infants with IUGR and same number of control infants Year of birth 1984-1985 Congenital anomaly NOT excluded Ethnicity unreported	Unreported	Population chart <10th centile (Singh chart)	1. Neonatal mortality (precise age unclear) 2. Neonatal hypoglycaemia ≤30mg/dL (2-20 hours of age)
Balcazar and Haas 1990 Mexico	Retrospective cohort	(9201) Singleton Infants born to women of low social class at a mother and child centre Year of birth 1981-1983 Congenital anomaly NOT excluded Ethnicity unreported	LMP	1.Population chart < 10th centile(Lubchenco chart) 2.Birth weight <2.9kg	Neonatal death (72 hours)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Berle et al 2006 Norway	Retrospective cohort	(7415) Singleton General population study, all those in age range in the country invited Year of birth 1967 onwards No exclusions specified Ethnicity unreported	LMP	Population chart <10th percentile (based on all live births in Norway 28-43 weeks)	1. Questionnaire: education level completed (primary school vs higher) (aged 20-30 years) 2.Questionnaire: socioeconomic functioning (receiving disability pension, un rehabilitation, being unemployed, or on sick leave) (aged 20-30 years) 3. Hospital Anxiety and Depression Rating Scale (HADS) cut off 8 (aged 20-30 years)
Bhargava, Sachdev and Ghosh 1985 India	Prospective cohort	(15596) Singleton Hospital population, consecutive births Year of birth unreported No exclusions specified Ethnicity unreported	LMP, confirmed by neonatal physical and neurological examination	1 Birth weight > -1 Standard Deviation (from population mean) 2. Birth weight> -2 Standard deviations (from population mean)	Neonatal mortality (7 days)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Bilge et al 2011 Turkey	Retrospective cohort	(59) Singleton Children born SGA at tertiary centre, with age-matched, non-obese controls Year of birth unreported Congenital anomaly excluded Ethnicity 100% Caucasian	LMP	Birth weight and /or length <-2 Standard Deviations for gestational age (unclear how standard derived)	1.Systolic blood pressure> 95th percentile for sex, age and height- single measurement (5.5-12.2 years old) 2. Diastolic blood pressure> 95th percentile for sex, age and height- single measurement(5.5-12.2 years old) 3. Systolic blood pressure> 95th percentile for sex, age and height-mean of 24 hour ambulatory monitoring (5.5-12.2 years old) 4. Diastolic blood pressure> 95th percentile for sex, age and height-mean of 24 hour ambulatory monitoring (5.5-12.2 years old
Burke et al 2004 Australia	Prospective cohort	(1913) Singleton Hospital population, offspring of women recruited during pregnancy Year of birth 1989 onwards Congenital anomaly excluded Ethnicity: 90% white, 5% Chinese, Indian 2%, Other 3%	Unclear	Birth weight < 2.5kg	Systolic blood pressure ≥ 120mmHg (8 years old)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Catalano et al 2009 USA	Prospective cohort	(89) Singleton Hospital population, offspring of pregnant women recruited with gestational diabetes or normal glucose tolerance Year of birth 1990-1999 Congenital anomaly excluded Ethnicity: 91% white, 3% African American, 4% Hispanic, 2% Asian	Unclear	Birth weight <2.5kg	1. BMI> 95th percentile (CDC criteria) (6-11 years) 2. Systolic blood pressure >95th percentile (reference unreported) (6-11 years) 3. Diastolic blood pressure > 95th percentile for (reference unreported) (6-11 years) 4. Serum cholesterol > 95th percentile (6-11 years)
Chaudhari et al 1996 India	Prospective cohort	(198) Hospital population, high risk infants (one or more of birth weight <2000g, Apgar <5 at 5 minutes or HIE, septicaemia or hyperbilirubinaemia, apnoea, seizures, IVH). Control infants of normal weight with uncomplicated antenatal and postnatal course Year of birth 1987-1992 Not specified if singleton/ multiple pregnancy Congenital anomalies excluded Ethnicity unreported	Unclear	Population chart <10th percentile (Singh criteria)	Cerebral palsy (definition unreported) (up to 12 months of age)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Cornelius et al 2009 USA	Prospective cohort	(1005) Hospital population: women who attended prenatal clinic, one group selected on the base of alcohol use (mixture of > or < 3 units per week, and women who used marijuana, and a random sample of those who did not. Also cohort of teenage mothers, from same clinic. Year of birth 1982-1985 Excluded if mother diabetic or child had sickle cell anaemia Singletons Ethnicity: 55% African American	Unclear	Absolute birth weight <2.5kg	BMI >95th percentile (CDC criteria) (age 6 years)
De Almeida and de Mello Jorge 1998 Brazil	Retrospective cohort	(2024) Hospital population (11 centres) Year of birth 1992 Singleton Excluded if gestational age unknown Ethnicity unreported	LMP	Population chart <10th percentile (Lubchenco chart)	Neonatal death (28 days)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Doctor et al 2001 USA	Retrospective cohort	(744) Hospital population Year of birth 1997 Mixed singleton and multiple pregnancy Infants with birth weight < 10th percentile, and matched controls birth weight 10th-90th percentile Excluded if gestational age unreliable Ethnicity: 38% Caucasian, 60% African American, 2% other	Second trimester ultrasound scan	Population chart <10th percentile (Alexander et al reference chart) Population chart <5th percentile (Alexander et al reference chart)	1.Neonatal intubation required (at birth) 2.Neonatal intensive care unit admission required (neonatal) 3.Respiratory distress (transient tachypnoea requiring oxygen RR>60/min for more than 4 hours/pneumonia clinical and radiological/meconium aspiration with supportive radiology (neonatal) 4.Hypoglycaemia (blood glucose < 40mg/dL and symptoms- jittery/ tachypnoeic/hypothermic) (neonatal) 5.Seizures (threshold unreported) (neonatal)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Evensen et al 2009 Norway	Prospective cohort	(142) Multiparous hospital population (multicentre) Year of birth 1986-1988 Cohort of birth weight <10th centile and random sample of birth weight ≥ 10th centile Excluded congenital anomaly or cerebral palsy Unclear if singleton or multiple pregnancy Ethnicity unreported	LMP + USS if discrepancy of more than 14 days or if could not be recalled accurately	Population chart <10th percentile (adjusted for parity, gestation and sex)	1. Visual acuity (<1.0 Snellen decimal) 2. Poor contrast sensitivity (Vistech chart, 1 or more values below normal) 3. Stereoacuity (above 240s of arc) 4. Strabismus (esodeviation/exodeviation larger than -8 at near, or -2PD at distance, any vertical deviations) 5. Nystagmus (pathological) 6. Accommodation (<6.5 diopters) 7. Convergence (near point convergence <10cm) 8. Visual perception (VMI-IV Developmental test of Visual Motor Integration test <22) 9. Overall visual abnormality (≥ 1 of above impairments) (age 14 years)
Evensen et al 2009 Norway	Prospective cohort	Same population as above study	LMP + USS if discrepancy of more than 14 days or if could not be recalled accurately	Population chart <10th percentile (adjusted for parity, gestation and sex)	Systolic blood pressure >140mmHg (age 18 years)
Evensen et al 2004 Norway	Prospective cohort	Same study population as above	LMP + USS if discrepancy of more than 14 days or if could not be recalled accurately	Population chart <10th percentile (adjusted for parity, gestation and sex)	1. Movement ABC score (total <14) (age 14 years) 2. Weschler IQ (> 2SD below control group mean) (age 14 years)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Ferdynus et al 2009 France	Retrospective cohort	(121312) Births in the Burgundy region from 2000 to 2006 Singleton Chromosomal anomalies excluded Ethnicity not reported	LMP and early USS	1. Population chart <10th percentile (based on whole study population) 2. 'Healthy population' chart <10th percentile (excluding maternal diabetes, hypertension, pre-eclampsia/eclampsia, abruption placentae, placenta praevia, presumed chorioamnionitis)	1. Neonatal mortality (death during hospital stay) 2. HIE (definition not specified) (neonatal)
Figueras et al 2007 Spain	Retrospective cohort	(12705) Hospital population Year of birth 2001-2005 Singleton Congenital and chromosomal anomalies excluded 58% multiparous mothers Ethnicity: 73.9% White, 4.6% South East Asian, 2.3% Central African, 18.2% South American, Other 0.9%	Early second trimester ultrasound	1. Population chart <10th percentile (Santamaria Spanish population chart) 2. Customised centile chart <10th percentile (Gardosi et al)	1. Neurological morbidity (Seizures, IVH> grade 2, PVL, HIE or abnormal EEG) (neonatal) 2. Non-neurological morbidity (NICU stay > 10 days, NEC (Walsh criteria), renal failure (SCr > 132.6 µmol/l / 1.5mg/dl), cardiac failure (need for inotropes)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Fitzhardinge and Steven 1972 Canada	Prospective cohort	(132) Hospital population Year of birth 1960-1966 Singleton Excluded major congenital anomalies, prenatal rubella infection, chromosomal anomalies, marked discrepancies in gestational age from LMP and neonatal findings Group of low birth weight infants and comparison groups of siblings of the low birth weight group who were of normal weight and birth history Ethnicity unreported	LMP, and neonates appearance and neurological development	30% under expected (Streeter's table, below 3rd percentile Stuart's table)	1. Speech defect (delayed onset, articulation defect, immature speech with poor receptive/ expressive ability or absent speech) (age 3-4 years) 2. CNS abnormality (including hyperactivity, short attention span, learning problems especially perceptive, poor fine co-ordination, hyper-reflexia, abnormal EEG) (age 5 years) 3. Convulsions (age 5 years) 4. Cerebral palsy (age 5 years) Specific thresholds for these conditions unreported
Gagliardo et al 2004 Brazil	Prospective cohort	(34) Hospital population, SGA neonate recruited followed by next 2 AGA neonates. All considered in good health for going home within 2 days of birth Year of birth 2000-2001 Excluded genetic syndromes, multiple anomalies, congenital infections Unclear if singleton or multiple pregnancies Ethnicity unreported	Capurro method	Population chart <10th percentile (Lubchenco) (AGA defined as 25th-90th percentile)	1. Bayley scale of motor development (BSD-II) <85 (age 3 months) 2. Bayley scale of mental development (BSD-II) <85 (age 3 months)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Gardner et al 2009 UK	Retrospective cohort	(242) Healthy children recruited from randomly selected schools, stratified for socioeconomic status Birth weight obtained from child health registry Year of birth >1994 Unclear if singleton or mixed No specific exclusions reported Ethnicity: 98% white	Unreported	Birth weight <2.5kg	1. Obese (>98th percentile UK 1990 growth charts) (age 9 years) 2. Overweight or obese (>91st percentile 1990 UK growth charts)(age 9 years)
Glinianaia et al 2010 UK	Retrospective cohort	(107461) All singleton births in Newcastle upon Tyne 1961-2000 Excluded individuals with missing data (congenital anomalies not excluded) Ethnicity unreported	Unreported	1. Birth weight <2.5kg 2. Birth weight <1.5kg	1. Neonatal death (28 days) 2. Infant death (≤ 12 months)
Gouyon et al 2003 France	Prospective cohort	(27008) General population, all live full term births in Burgundy region Year of birth 2000-2001 Ethnicity unreported Not specified if singleton/ multiple births No exclusions reported	Unreported	1. Population chart <10th percentile (French growth curves) 2. Population chart <3rd percentile (French growth curves)	1. Severe neurological disorders (threshold unreported) (neonatal) 2. Respiratory distress syndrome (threshold unreported) (neonatal) 3. Neonatal death

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Gray et al 1991[81] Brazil	Prospective cohort	(2977) Hospital and maternity unit population (5 centres) All low birth weight infants and control group (selection method unclear) Year of birth 1984-1986 Singleton Congenital anomalies not excluded Ethnicity unreported	Capurro method	Birth weight <2.5kg	1. Early neonatal death (7days)
Haas, Balcazar and Caulfield 1987 Mexico	Retrospective cohort	(16321) 2 maternity hospitals, one serving lower socioeconomic class, one mixed population. Year of birth 1981-1984 Excluded outliers <500g or > 5500g, length <40 cm or > 56cm. Ethnicity Latin American	Classified by neonatal clinical examination as term or preterm, LMP addition if available	1.Population chart <10th percentile (Brandt) 2.Rohrer's Index (wt (g) x100/length (cm)3)	Early neonatal death (48 hours)
Hands et al 2009 Australia	Prospective cohort	(1555) Hospital population, women recruited from tertiary centre antenatal clinic Year of birth 1989 onwards Singleton Congenital anomaly excluded Ethnicity: 84.9% Caucasian, 2.1 % Aboriginal, 12.1% Other	Unreported	Birth weight <2.0kg	Mild motor disability (Neuromuscular development index (NDI) from scores of McCarron Assessment of neuromuscular development (MAND) mild disability 71-85, >85 normal) (age 10 years)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Hemachandra et al 2006 USA	Retrospective cohort	(24951) Pregnant women enrolled at 12 academic medical centres Year of birth 1959-1965 Mixed singleton and multiple births Exclusions: Deaths 0-7 years of age, implausible data (≥ 4SD from mean), children diagnosed with cardiac or renal disease Ethnicity: 52% white and 48% black	LMP	Birth weight ≤ 2.5kg	Systolic BP>90th percentile (study population blood pressure distributions) (age 7 years)
Hindmarsh et al 2010 UK	Prospective cohort	(339) Pregnant women enrolled at tertiary referral centre Year of birth 1996-1997 Singleton Excluded if congenital anomalies or maternal steroid use Caucasian mothers	LMP, corrected by USS if CRL <12 weeks or BPD 12-20 weeks differed from LMP date by >7 days	1.Birth weight <2.5kg 2. Birth weight <1.5kg	1. Systolic BP >90th centile (4th task force on blood pressure in children) (age 3 years) 2. Diastolic BP >90th centile(4th task force on blood pressure in children) (age 3 years)
Indredavik et al 2010 Norway	Prospective cohort	(140) Year of birth 1986-1988 Cohort of birth weight <10th centile and random sample of birth weight ≥ 10th centile Excluded congenital anomaly or cerebral palsy Unclear if singleton or multiple pregnancy Ethnicity unreported	LMP + USS if discrepancy of more than 14 days or if could not be recalled accurately	Population chart <10th percentile (Vik definition, percentiles adjusted for gestational age, sex and parity)	Psychiatric diagnosis (DSM-IV anxiety disorder, Attention Deficit Hyperactivity disorder, Autism Spectrum Disorder) (age 14 years)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Jacobsson et al 2008 Sweden	Case control study	(654) Cases from cerebral palsy register, born in Sweden and had lived in the study area at 4-8 years of life. Matched with 2 controls from national birth register, closest births to index case matched for gestational age, gender, age, sex, multiple pregnancy and delivery ward Year of birth 1983-1990 Singleton Excluded if postnatal cause of cerebral palsy or incomplete data Ethnicity unreported	USS between 16 and 19 weeks 97% cases, 3% LMP only.	1. Customised chart <10th percentile (Gardosi) 2. Customised chart <5rd percentile (Gardosi) 3. Customised chart <1st percentile (Gardosi)	Cerebral palsy (Swedish classification non progressive motor impairment) (4-8 years old)
Jeliffe- Pawlowski and Hansen 2004 USA	Retrospective cohort	(27791) Pregnant women enrolled at 12 academic medical centres Year of birth 1959-1965 Singleton Exclusions: Deaths 0-7 years of age, implausible data (≥ 4SD from mean), born after 44 weeks gestation Ethnicity: 50% white and 50% black	LMP	Population chart <10th percentile (gender and weight specific for this cohort) (AGA= 25thto 90th percentile)	1. Bayley mental development index (MDI) score ≤85(age 8 months) 2. Bayley psychomotor development index score(PDI) ≤85(age 8 months) 3. Combined delay (MDI and PDI≤85) (age 8 months) 4. Mental retardation (Stamford-Binet IQ <70) (age 4years)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Johnson 1985 USA	Case control study	(303) Cases of neuroblastoma and random sample of controls born in Texas Unclear if singleton/ mixed Year of birth 1964 to 1978 No exclusions specified Ethnicity: mixed White, Hispanic and Black population	LMP	Birth weight <2.5kg	Death from neuroblastoma (death certificate cause of death) (age before age 15 years)
Kajante et al 2006 Finland	Retrospective cohort	(13830) Individuals born in University hospital, went to school in Helsinki and still resident in Finland at time of study. Year of birth 1924-1944 Singleton No exclusions specified Ethnicity unreported	LMP	Birth weight <2.5kg	1. All cause mortality (Finnish death register) (26.7-74.9 years) 2. All cause mortality (<55 years of age) 3. Cancer mortality (26.7-74.9 years) 4. Cardiovascular mortality (26.7-74.9 years)
Kindlund et al 2010 Denmark	Retrospective cohort	(4989) Live born twins in Denmark during study period who were included in nationwide Danish Twin Registry Year of Birth 1994-2000 No other exclusions specified Ethnicity unreported	Unreported	Birth weight <2.5kg	Asthma (questionnaire to parents about diagnosis) (age 3-9 years)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

KlungsØyr Melve and Skjaerven 2003 Norway	Retrospective cohort	(527157) Births in the Norwegian Medical Birth Registry, mothers with at least 2 singletons Year of Birth 1967 to 1998 Singleton No exclusions specified Ethnicity unreported	LMP	1.Birth weight <2.5kg 2.Birth weight <2.0kg 3.Birth weight <1.5kg	Neonatal death (28 days)
Kotecha et al 2010 UK	Retrospective cohort	(5770) Population study open to pregnant women residing in Avon Year of birth 1991-1992 Singleton Not specified if anomalies excluded Ethnicity: 100% Caucasian	LMP, confirmed by scan in a proportion of the population	Z score (birth weight and, adjusted for gestational age and gender) < 1.28 (10th centile)	Asthma (American thoracic society standards) (age 91 months)
Kramer et al 1990 Canada	Retrospective cohort	(5305) Hospital born infants at single centre. Year of birth 1980-1986 Singleton Excluded congenital anomalies, chromosomal anomalies and evidence of intrauterine infection Ethnicity unreported	USS BPD 16-18 weeks gestation + LMP	1.Fetal growth ratio <0.75 (observed birth weight/ mean birth weight for gestational age for this hospital population) 2. Fetal growth ratio <0.80(observed birth weight/ mean birth weight for gestational age for this hospital population)	1.Neonatal death (in hospital, age range unspecified) 2. Hypoglycaemia (plasma glucose <40mg/dL) (neonatal)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Kuhle, Allen and Veugelers 2010 Canada	Retrospective cohort	(3190) Population based study, Grade 5 students in Nova Scotia region, linked to perinatal database. Year of birth: 1991-1992 Unclear if singleton/ multiple births Exclusions missing or improbable birth weight for gestational age, invalid health insurance number Ethnicity unreported	Unreported	Population chart <10th percentile (Kramer definition)	Overweight or obese (CDC classification) (age 11 years)
Lamb et al 2010 USA	Retrospective cohort	(1178) Two groups of children: one recruited in infancy (unaffected first degree relatives of patients with type 1 diabetes, identified through childhood diabetes centre; second group babies born at single hospital in Denver region. Year of birth: unreported Unclear if singleton/ multiple births Ethnicity: non-Hispanic white 72.1%	Unreported	Population chart <10th percentile (Oken USA reference chart)	1. BMI ≥ 85th percentile (CDC growth charts) (age 2-11.5 years) 2. BMI ≥ 95th percentile (CDC growth charts) (age 2-11.5 years)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Larroque et al 2001 France	Retrospective cohort	(496) Population based birth registry in and around the city of Haguenau, all full term singletons born SGA, control first AGA infant listed in registry after SGA infant Year of birth 1971 to 1978 Singleton Excluded aberrant growth measurements, psychomotor disorder or institutionalised, or chronic illness including congenital anomaly, adoption, death Ethnicity unreported	LMP, physical examination, confirmed by USS where available	Population chart <3rd percentile birth weight or length (Haguenau local reference curves)	School difficulties (Late entry into secondary school) (age 11-12 years)
Levitt et al 2000 South Africa	Prospective cohort	(113) Single hospital population, primigravid mothers. All infants born SGA included, and random sample of AGA Year of birth 1975-1976 Singleton No exclusions reported Ethnicity: mixed Koi San, European, East Indian, Malaysian and Black African	Dubowitz method	Population chart <10th centile, compared with individuals 25th-75th percentile	1. Impaired glucose tolerance and type 2 diabetes (1985 WHO criteria) (age 20 years) 2. Hypertension (BP>140/90mmHg) (age 20 years)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Li, Law and Power 2007 UK	Retrospective cohort	(8520) Infants born in England, Wales and Scotland in one week period Year of birth 1958 Not specified if singleton/ multiple births No exclusions reported Ethnicity unreported	Unreported	Birth weight <2.5kg	1. Hypertension (BP ≥ 140/90mmHg) (age 45 years) 2. BMI ≥ 30 (age 45 years)
Libby et al 2008 UK	Retrospective cohort	(1065) Population study: Dundee birth cohort, record linked to adult health factors in Scottish working population Year of birth 1952-1966 Unclear if singleton/ multiple births No exclusions reported Ethnicity unreported	LMP	Birth weight <2.5kg	Total cholesterol >5mmol/L (non-fasting) (age 24 to 42 years)
Liew et al 2008 USA	Retrospective cohort	(609) Population based cohort, adults recruited in middle age by probability sampling in 4 US communities Year of birth 1923-1944 Unclear if singleton/ mixed Excluded if retinopathy not classifiable Ethnicity: 79% White	Self- report of prematurity	Birth weight >2.5kg	Diabetic retinopathy any grade (American academy of opthalmologists 2007)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Lira, Ashworth and Morris 1996 Brazil	Prospective cohort	(393) Infants recruited from the maternity wards of two state hospitals, one private hospital and three government health centers. Eligible if 1500-2499g, controls of weight 3000-3499g matched for sex and season of birth Year of birth 1993-1994 Congenital anomalies excluded Singleton Ethnically diverse (not specified)	Capurro method	Birth weight >2.5kg	Death (up to 26 weeks of age)
Low et al 1992 Canada	Prospective cohort	(104) Cohort of high risk infants, recruited if FGR <10th centile, infants of IDDM mothers, intrapartum asphyxia with um artery buffer base < 34mmol/L, respiratory complications or encephalopathy Year of birth 1978-1982 Mixed singleton and multiple births Excluded congenital or genetic anomaly Ethnicity unreported	Unclear	Population chart <10th centile (reference chart unreported)	Learning deficit (< 15th percentile than expected for age Woodcock reading mastery test) (age 9-11 years)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Low et al 1982 Canada	Prospective cohort	(124) Group of IUGR infants (birth weight <10th percentile) and group of control infants (birth weight >25th percentile) Year of birth: unreported Mixed singleton/multiple birth Ethnicity:96% of control infants and 86% of IUGR infants Caucasian	LMP	1. Population chart <10th percentile (Gruenwald curve) 2. Population chart <5th percentile (Gruenwald curve)	1. Intelligence (Weschler Pre-school and primary scale of Intelligence <85) (age 60 months) 2. Motor handicap (minor abnormalities of tone, co-ordination or motor function or McCarthy's motor score ≤39, or cerebral palsy) (age 60 months)
Lubchenco, Searls and Brazie 1972 USA	Retrospective cohort	(11197) All live born infants at single university centre during the study period, with adequate information on gestational age Year of birth: 1958 to 1968 Mixed singleton/ multiple birth No exclusions reported Ethnicity 55% Anglo-American, 30% Mexican American, 15% African American	Unreported	1.Birth weight <2.5kg 2.Birth weight <2.0kg	Neonatal death (28days)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Lurbe et al 2009 Spain	Case control study	(422) Obese adolescents attending Paediatric unit, non- obese subjects drawn from a parallel study, all well and not taking medication. Birth details obtained from obstetric records. All mothers normotensive during pregnancy Year of birth: unclear Excluded if: severe obesity, secondary obesity Unclear if singleton/ multiple birth Ethnicity: all white European	Unreported	Population chart <10th percentile (Lubchenco)	Obesity (BMI >97th percentile for age and sex) (age 10-18 years)
Manji, Massawe and Mgone 1998 Tanzania	Prospective cohort	(800) Neonatal unit at teaching hospital, admissions over a 4 months period Year of birth 1990-1991 Exclusions unreported Mixed singleton/ multiple births Ethnicity unreported	Dubowitz method	1.Birth weight < 2.5kg 2.Birth weight <2.0kg 3.Birth weight <1.5kg	Death (up to 6 weeks of age)
Mardones et al 2008 Chile	Retrospective cohort	(216315) Live births and neonatal death information from Chilean Civil Registry service Year of birth: 2000 Mixed singleton and multiple births Excluded if data incomplete Ethnicity unreported	LMP, USS before 20 weeks, postnatal clinical examination. Majority have USS dating but exact no not reported	1. Birth weight <2.5kg 2. Birth weight <1.5kg	Neonatal death (28 days)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Matijasevich et al 2008 Brazil	Prospective cohort	(4130, 4171, 3392)Three separate birth cohorts in city of Pelotas. Maternity wards visited on daily basis and mothers interviewed. Perinatal mortality surveillance: visiting maternity wards, intermediate care centres, registry offices and cemeteries Year of birth : 1982, 1993, 2004 Mixed singleton and multiple births Excluded if gestational age incompatible with birth weight. Congenital anomalies not excluded. Ethnicity unreported	LMP, in 1993 and 2004 Dubowitz score also used, in 2004 large number also had early pregnancy USS	Population chart < 10th percentile (Williams curve)	Early neonatal death (7 days)
McIntire et al 1999 USA	Retrospective cohort	(82361) Live births at single hospital during study period. Outcomes abstracted from medical records, birth information prospective database. Year of birth 1988-1996 Singleton Congenital anomalies excluded, and birth weight >75th centile Ethnicity: 54% Hispanic, 28% black, 15% white, 3% other	LMP, if fundal height discrepancy (18-30 weeks gestation) > 2cm confirmed by USS	1. Population chart ≤ 10th percentile (this population) 2. Population chart ≤ 5th percentile (this population) 3. Population chart ≤ 3rd percentile (this population)	1. Neonatal death (28 days) 2. Seizures (definition unreported)(1st 24 hours)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

McKinney et al 1999 UK	Case control study	(598) Cases registered to York Childhood diabetes Register, two matched controls (year and month of birth) randomly selected from primary care registrations of FHSA Year of Birth: 1978-1994 Mixed singleton and multiple births Congenital malformations not excluded Ethnicity: 95% White, 3.5% Indian/Pakistani	Unreported	Birth weight <2.5kg	Type 1 Diabetes (diagnostic criteria unreported) (age 0-16 years)
Meas et al 2010 France	Retrospective cohort	(1308) Population based registry from Haguenau, SGA group individuals born <10th percentile, AGA group 25th -75th percentile Year of birth: 1971 - 1985 Singleton Excluded chronic disease including metabolic disorders Ethnicity unreported	Unreported	Population chart (<10th percentile) (Local growth curves)	1.Metabolic syndrome (at least 3 of: fasting blood glucose ≥ 6.1mmol/l, waist circ ≥ 102 cm (m) 88cm (f), triacylglycerol ≥1.69mmol/l, HDL cholesterol <1.04mmol/l (m) or (1.29mmol/l (f) and BP ≥130/85mmHg or treated hypertension) (age30 years) 2. Diabetes mellitus (criteria unreported) (age 30 years) 3. Diabetes mellitus or impaired glucose tolerance (OGTT, threshold unreported) (age 30 years)
Meas et al 2008 France	Retrospective cohort	As above	Unreported	Population chart <10th percentile (Local growth curves)	BMI >30 (age 30 years)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Menezes et al 2006 Brazil	Prospective cohort	(3931) All hospital deliveries in Pelotas region Year of birth 1993 Unclear if singleton/ multiple births No exclusions specified Ethnicity unreported	Dubowitz method	Birth weight <2.5kg	Systolic BP ≥ 120mmHg (1. Age 11 years 2. Age 15 years)
Mi et al 2008 China	Retrospective cohort	(932) Sequential live births at single hospital during the study period. Infants traced in adulthood Year of birth 1948-1954 Singleton No exclusions specified Ethnicity unreported	LMP	1.Birth weight <2.5kg 2. Ponderal Index (kg/m^3)<24.5	1. Metabolic syndrome (Alberti et al International Diabetes Federation definition) 2. Total cholesterol >5mmol/L 3. BMI >30 4. Diabetes mellitus type 2 (WHO criteria) 5. Diabetes mellitus or impaired glucose tolerance (OGTT,2hr glucose>7.8mmol/L) 6.Hypertension (pre-existing diagnosis) 7. Systolic BP ≥140mmHg 8. Diastolic BP ≥90mmHg All above at age 41-52 years
Minikami, Izumi and Sato 1999 Japan	Retrospective cohort	(54802) Japanese Ministry of Health and Welfare data regarding births and deaths Year of birth: 1989-1993 All multiple births No exclusions specified Ethnicity unreported	Exact numbers unreported: "early USS widely practiced in Japan/ many multiple pregnancies conceived after IVF"	Birth weight > -1SD below mean (calculated from study population)	Early neonatal death (7days)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Minior and Divon 1998 USA	Retrospective cohort	(268) Cohort of SGA infants born following uncomplicated pregnancies (excluded maternal medical conditions e.g. pre-eclampsia, diabetes, renal or automimmune disease), matched to three AGA controls Year of birth 1988-1995 Singleton Congenital or chromosomal anomalies excluded Ethnicity 35% white, 31% Black, 27% Hispanic, 7% Other	Early USS	1.Population chart <10th percentile (Brenner) 2.Population chart <5th percentile (Brenner)	Neonatal morbidity (one of more of: hypoglycaemia, respiratory distress, thrombocytopenia, hyperbilirubinaemia, intubation, sepsis, IVH, apnoea) (thresholds unreported)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Moser et al 2007 UK	Retrospective cohort	(590430) Routine birth and death registration records, and NHS Numbers for Babies data for babies born in England and Wales Year of Birth: 2005 Mixed singleton/ multiple birth Excluded if incomplete data Ethnicity unreported	Unreported	1.Birth weight <2.5kg 2.Birth weight <1.5kg	1. Neonatal death (up to 28 days) 2. Infant death (up to 12 months)
Nelson and Broman 1977 USA	Prospective cohort	(29551) Pregnant women enrolled at 12 academic medical centres Year of birth 1959-1965 Mixed singleton and multiple births Exclusions: children with gross CNS malformation, metabolic, chromosomal or other specific disorders or known neurological catastrophe e.g. meningitis post neonatal discharge Ethnicity: 52% white and 48% black	LMP	Absolute birth weight ≤ 2.5kg	Severe mental or motor handicap (IQ < 50 and moderate or severe motor deficits due to cerebral palsy) (age 7 years)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Nobili et al 2007 Italy	Case control study	(180) Cases of children with non-alcoholic fatty liver disease, consecutively observed in Liver Unit, and matched controls (age and sex) with normal liver scan and liver function Year of Birth: 1984-1998 Exclusions unreported Unreported if singleton/multiple birth Ethnicity unreported	Unreported	Population chart ≤10th percentile (Gairdner- sex and age specific)	Non-acute fatty liver disease (liver biopsy diagnosis) (mean age 11.3+/-3.8)
North 1966 USA	Retrospective cohort	(2676) Deliveries at single maternity unit, infants weighing 1.0-2.5kg matched with infants 3.0-3.5kg. Year of birth 1957-1964 Mixed singleton/multiple birth No exclusions specified Ethnicity: 75% White, 25% Non-white	Unreported	Birth weight ≤2.5kg	Neonatal death

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

O'Keefe et al 2003 Australia	Prospective cohort	(5023) Pregnant women recruited from single centre. Infants followed up. Year of birth: 1981-1984 Singleton Exclusions: perinatal death and adoption Ethnicity: 97% White	Unreported	1. Population chart <10th percentile (Roberts Australian norms, sex and gestation specific) 2. Population chart <3rd percentile (Roberts Australian norms, sex and gestation specific) 3.Ponderal index (kg/m3)< 25th percentile 4.BHR (birth weight (g)/ head circumference (cm)) <25th percentile	1. Learning difficulties (questionnaire to mother) (age 14 years) 2. WRAT reading score <85 (age 14 years)
Ott 1995 USA	Retrospective cohort	(957) Infants of a series of high risk obstetric patients referred to single perinatal centre. Year of birth: 1990-1991 Unclear if singleton/ multiple birth Congenital anomalies excluded Ethnicity unreported	Unreported	Population chart <10th percentile (Ott and Hadlock)	Neonatal mortality
Owen et al 2003 UK	Retrospective cohort	(1319) Random selection of children from 68 British secondary schools, birth records obtained from hospital of birth Year of birth:1982-1986 Mixed singleton/ multiple birth No exclusions reported Ethnicity: 92% White, 6% South Asian, 2% Other	Unreported	Birth weight <2.5kg	Total cholesterol <5mmol/L (fasting/non-fasting) (age 13-16 years)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Padin-Rojas et al 1986 Puerto Rico	Prospective cohort	(48) Cohort of infants born at tertiary centre and recruited to trial of infant nutrition Year of birth: 1986 Unclear if singleton/ multiple birth No exclusions specified Ethnicity unreported	LMP	Ponderal Index weight (g) x100/length $(cm)^3$ (low, threshold unspecified)	Bayley Mental Development Index <69 (age 1 year)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Pandolfi et al 2008 Italy	Retrospective cohort	(50) Women who had birth weight <2.5kg randomly recruited from neonatal unit registry of single hospital. Control subjects, next female in the registry with birth weight ≥ 3.0kg Year of birth: >1980 Singleton Excluded if: Congenital malformations, chromosomal anomalies, major neonatal and maternal pregnancy complications including gestational diabetes and pre-eclampsia Ethnicity unreported	Unreported	Birth weight <2.5kg	Polycystic ovarian syndrome (12 or more follicles < 10mm diameter or increased ovarian vol > 10cm3, associated with irregular menses, hirsutism) (age 21.8 +/- 1.4 years)
Pathai, Cumberland and Rahi 2010 UK	Retrospective cohort	(13613) Cohort of infants resident in UK, aged 9 months at time of recruitment. Details regarding birth weight and outcomes obtained through questionnaire to parents. Year of birth:2000 Mixed singleton/ multiple birth No exclusions specified Ethnicity: 89% White, 6% South Asian, 2.5 % Black African/Caribbean	Unreported	Birth weight <2.5kg	Strabismus (questionnaire to parents about strabismus and other eye conditions, common synonyms for strabismus considered positive response) (age 3 years)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Paz et al 1995 Israel	Retrospective cohort	(1707) All subjects born at University Medical center and subsequently drafted into the army. Year of birth: 1970 to 1971 Singleton Congenital anomalies, chromosomal disorders or congenital infections excluded Ethnicity: 11% Israeli, 21% Asian, 42% Euro-American, 26% African	LMP	Population chart <3rd percentile (based on study population)	Low educational achievement (less than 12 years of schooling or attending vocational school) (age 18 years)
Pearce and O'Sullivan 2003 UK	Retrospective cohort	(408) Children recruited from 26 schools in Newcastle upon Tyne area, birth information obtained from parents, confirmed in >50% by hospital records Year of birth: unclear Unclear if singleton/ multiple birth Excluded pre-existing cardiac or renal disease Ethnicity: unreported	Unreported	Birth weight <2.5kg	1. Systolic BP >120mmHg (age 6-16years) 2.Diastolic BP >80mmHg (age 6-16 years)
Peng et al 2005 China	Retrospective cohort	(147) Infants born at two Shangai hospitals. Year of birth 1983 Mixed singleton/ multiple births Excluded if birth weight <1.2kg Ethnicity: Chinese	LMP and newborn physical examination	Population chart <10th percentile (chart unreported)	1. Neurodevelopmental abnormalities (Gesell Developmental diagnosis DQ ≤ 85)(1. Age 4 months 2. Age 36 months) 2. IQ (Weschler IQ ≤85) (Age 16 years)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Plante 2004 USA	Retrospective cohort	(7802) All females born at term in Pennsylvania in 1974 who delivered full term singleton live born infants 1999-2000, data regarding outcome obtained from check box on birth certificate Year of birth:1974 Singleton Excluded racial origin other than black or white, or incomplete information Ethnicity: 79% White 21% Black	clinical, LMP or ultrasound criteria	Population <10th centile (Brenner chart, gestation and race specific)	1. Diabetes in pregnancy (gestational/ pre-existing)
Pulver et al 2009 USA	Retrospective cohort	(316077) Population study of infants born in Utah, linked birth and death databases Year of birth: 1999-2005 Unclear if singleton/ mixed Congenital anomalies not excluded Ethnicity unreported	LMP, prenatal USS and/or newborn examination	Population <10th percentile (curves for the study population)	1.Neonatal death (up to 28 days) 2.Infant death
Rahaila et al 2002 Finland	Retrospective cohort	(100) All children with IUGR born at single university hospital, next AGA child of same gender and born at term selected as control. Year of birth: 1984 to 1986 Singleton Congenital anomalies excluded Ethnicity unreported	Unreported	>2SD below mean weight for gestational age	ABPM, mean systolic BP >120mmHg (age 12 years)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Rich-Edwards et al 1997 USA	Retrospective cohort	(66689) Nurses' health study, postal questionnaires sent to registered nurses regarding their medical histories and lifestyle. Birth information self- reported. Outcome data from medical records Year of birth: 1921-1946 Singleton Excluded if had cardiovascular disease at time of initial questionnaire, or birth weight unknown Ethnicity: unreported	Self-report of prematurity, method of diagnosis unclear	Birth weight ≤2495g	1. Non-fatal coronary heart disease (WHO criteria: symptoms and either diagnostic ECG or cardiac enzyme changes)(age 46-71 years) 2. Non-fatal stroke (typical neurological deficit of sudden onset for more than 24 hours. Ischaemic and haemorrhagic) (age 46-71 years)
Salonen et al 2010 Finland	Prospective cohort	(70) All children with IUGR born at single university hospital, next AGA child of same gender and born at term selected as control. (AGA described as ≤ -2SD and ≤ +2SD scores of the respective mean for gestational age and sex) Year of birth: 1984 to 1986 Singleton Congenital anomalies excluded Ethnicity unreported	Unreported	>2SD below mean weight for gestational age	1. Systolic BP >140mmHg (age 20 years) 2. Total cholesterol >5mmol/L (after overnight fast)(age 20 years) 3. BMI ≥ 30(age 20 years)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Sommerfelt et al 2001 Norway	Prospective cohort	(625) Multicenter study, parous mothers recruited before 20 weeks gestation. Those with risk factors for SGA infant and 10% random sample of the rest followed up Year of birth 1986 to 1988 Unclear if singleton/ multiple births Congenital anomalies excluded, and if unable to speak Scandinavian language Ethnicity unreported	USS 17-18 wks, altered date from LMP if uncertain or >14 days discrepancy	Population chart <15th percentile (reference standards from Norwegian Birth Registry)	Behavioural problems (total problem score (YCI and ERS), cut of > 95th percentile for scores from AGA group)(age 5 years)
Strauss 2000 UK	Prospective cohort	(3034) Infants born in the UK during a single week, followed up with questionnaires and academic tests Year of birth: 1970 Mixed singleton/ multiple birth Congenital anomalies excluded Ethnicity: 97% White British/European	LMP	Population chart <5th percentile (British reference standards (Cole/ Davie))	Impaired academic ability (teachers assessment bottom 15th percentile)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Thompson et al 2001 UK	Retrospective cohort	(810) Subsample of study of all births in Herefordshire, traced and still living in the county. Year of birth:1920-1930 Singleton No exclusions specified Ethnicity unreported	Unreported	Birth weight <2.9kg	Depression (5 or more Geriatric Depression Score, or 3 or more Geriatric Mental State B version) (mean age 68 years)
Ullah et al 2009 Bangladesh	Prospective cohort	(770) Term newborns at a single hospital, stratified sampling from three birth weight strata (≤2, >2 to <2.5 and ≥2.5kg) Year of birth: unreported Singleton Included infants defined as normal, therefore anomalies excluded Ethnicity: Bangladeshi	LMP	1. Birth weight <2.5kg 2. Birth weight ≤ 2kg	1. Neonatal death (1st 7 days) 2. Birth asphyxia (definition not described)(1st 7 days) 3. Acute respiratory infection (definition not described) (1st 7 days)
Uvebrant and Hagberg 1992 Sweden	Case control study	(613) Population based series of children with cerebral palsy and controls, term live born infant of the same sex and born in the same unit Year of birth: 1975-1982 Mixed singleton/ multiple birth Excluded: missing data or postnatal (after first week of life) cause of cerebral palsy Ethnicity: unreported	Unreported	≥-2SD below mean (Fryer Swedish growth chart)	Cerebral palsy (non-progressive disorder movement/ posture, Bax definition) (childhood, age not specified)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Vikse et al 2008 Norway	Retrospective cohort	(1937768) Birth data from Norwegian birth registry matched to data from Norwegian renal registry Year of birth: 1967 onwards Mixed singleton/ multiple birth Congenital anomalies not excluded Ethnicity unreported	Unreported	Population chart <10th percentile	End stage renal disease(definition not reported) (0.2-38 years)
Walther and Ramaekers 1982 The Netherlands	Cohort, unclear if prospective/ retrospective	(50) 25 consecutively born term infants showing intrauterine malnutrition prospectively compared to 25 normally grown term infants Year of birth: not reported Singleton Congenital anomalies excluded Ethnicity: all Caucasian	Dubowitz method	Population chart <10th percentile (Kloosterman grid)	Delayed language achievement (Reynell Developmental Language scales ≥2 SD below standard mean) (age 31-42 months)
Walther and Ramaekers 1982 The Netherlands	Cohort, unclear if prospective/ retrospective	(500) consecutive live born infants admitted to the neonatal ward directly after birth at a single hospital. Year of birth: Unreported Singleton Congenital anomalies, chromosomal anomalies, fetal infections or haemolytic disease excluded. Ethnicity: all Caucasian	Dubowitz method	1.Population chart <10th percentile (Kloosterman grid) 2. Ponderal Index (kg/m^3) <10th percentile (Miller and Hassanein chart)	1. Hypothermia (admission rectal temp < 35.5 C)(neonatal) 2. Hypoglycaemia (≤ 1.6mmol/L blood glucose)(neonatal)

Appendix 11. Characteristics of included studies in systematic review of birth weight standards

Wei et al 2006 Taiwan	Case control study	(810)All school children in Taiwan province screened for diabetes, all cases of type 1 1992-1997 and random selection of normal controls. Birth data obtained from Birth Registry Year of birth:1974 to 1991 Unclear if singleton/multiple birth Excluded if missing data Ethnicity: unreported	Unreported	Birth weight ≤2.6kg	1.Type 1 Diabetes mellitus (≥ 126mg/dl fasting blood glucose, received insulin within 6 months, or diagnosis by referring physicians) (6-18 years) 2. BMI (≥95th percentile for sex and age specific anthropometrics of children in Taiwan) (6-18 years)
Wennergren 1986 Sweden	Prospective cohort	(4415) Unselected population, all infants born in Goteburg Year of birth:1978 Mixed singleton/ multiple births Exclusions: missing data Ethnicity: unreported	Unreported	Birth weight ≥-2SD below mean (standard unreported)	Respiratory disorders (Respiratory rate ≥ 60/ min or <30/ min during first 3 hours; grunting/ chest wall recessions at 2hrs age or later; central cyanosis or apnoeic spells) (neonatal)
Zhang et al 2007 Sweden	Retrospective cohort	(751281) Population based, data from Swedish birth registry Year of birth 1992-2001 Singleton Exclusions: missing data Ethnicity: 87% Nordic	USS 18 weeks gestation	Population chart <10th percentile (centiles based on the study population)	Early neonatal death (first 7 days)

Appendix 12. Reference list of included studies in systematic review of birth weight standards

Algert CS, McElduff A, Morris JM, Roberts CL. Perinatal risk factors for early onset of Type 1 diabetes in a 2000-2005 birth cohort. *Diabetic Medicine* 2009; 26(12):1193-1197.

Als H, Tronick E, Adamson L. The behaviour of the full term but underweight newborn infant. *Developmental Medicine and Child Neurology* 1976; 18:590-602.

Amigo H, Bustos P, Alvarado ME, Barbieri M, Bettiol H, da Silva AAM et al. Size at birth and lipoprotein concentrations in adulthood: two prospective studies in Latin American cities. *Journal of Epidemiology and Community Health* 2010; 64(10):855-859.

Ananth CV, Vintzileos AM. Distinguishing pathological from constitutional small for gestational age births in population-based studies. *Early Human Development* 2009; 85(10):653-658.

Andersson SWL. Blood pressure and hypertension in middle-aged women in relation to weight and length at birth: A follow-up study. *Journal of Hypertension* 2000; 18(12):2000.

Arora NK, Paul VK, Singh M. Morbidity and mortality in term infants with intrauterine growth retardation. *Journal of Tropical Pediatrics* 1987; 33:186-189.

Balcazar HH. Classification schemes of small-for-gestational age and type of intrauterine growth retardation and its implications to early neonatal mortality. *Early Human Development* 1990; 24(3):1990.

Berle JO, Mykletun A, Daltveit AK, Rasmussen S, Dahl AA. Outcomes in adulthood for children with foetal growth retardation. A linkage study from the Nord-Trondelag Health Study (HUNT) and the Medical Birth Registry of Norway. *Acta Psychiatrica Scandinavica* 2006; 113(6):501-509.

Bhargava SK, Sachdev HP, Ghosh S, Bhargava SK, Sachdev HP, Ghosh S. Distribution of live births & early neonatal mortality in relation to gestation & intrauterine growth. *Indian Journal of Medical Research* 1985; 82:95-97.

Bilge I, Poyrazoglu S, Bas F, Emre S, Sirin A, Gokalp S et al. Ambulatory blood pressure monitoring and renal functions in term small-for-gestational age children. *Pediatric Nephrology* 2011; 26(1):119-126.

Burke V, Beilin LJ, Blake KV, Doherty D, Kendall GE, Newnham JP et al. Indicators of fetal growth do not independently predict blood pressure in 8-year-old Australians - A prospective cohort study. *Hypertension* 2004; 43(2):208-213.

Catalano PM, Farrell K, Thomas A, Huston-Presley L, Mencin P, de Mouzon SH et al. Perinatal risk factors for childhood obesity and

metabolic dysregulation. *American Journal of Clinical Nutrition* 2009; 90(5):1303-1313.

Chaudhari S, Kulkarni S, Barve S, Pandit AN, Sonak U, Sarpotdar N et al. Neurologic sequelae in high risk infants--a three year follow up. *Indian Pediatrics* 1996; 33(8):645-653.

Cornelius MD, Goldschmidt L, Willford JA, Leech SL, Larkby C, Day NL. Body Size and Intelligence in 6-year-olds: Are Offspring of Teenage Mothers at Risk? *Maternal and Child Health Journal* 2009; 13(6):847-856.

de Almeida MF, Jorge MHPD. Small for gestational age: risk factor for neonatal mortality. *Revista de Saude Publica* 1998; 32(3):217-224.

Doctor BA, O'Riordan MA, Kirchner HL, Shah D, Hack M, Doctor BA et al. Perinatal correlates and neonatal outcomes of small for gestational age infants born at term gestation. *American Journal of Obstetrics & Gynecology* 2001; 185(3):652-659.

Evensen KA, Vik T, Helbostad J, Indredavik MS, Kulseng S, Brubakk AM et al. Motor skills in adolescents with low birth weight. *Archives of Disease in Childhood Fetal & Neonatal Edition* 2004; 89(5):F451-F455.

Evensen KA, Lindqvist S, Indredavik MS, Skranes J, Brubakk AM, Vik T et al. Do visual impairments affect risk of motor problems in preterm and term low birth weight adolescents? *European Journal of Paediatric Neurology* 2009; 13(1):47-56.

Evensen KAI, Steinshamn S, Tjonna AE, Stolen T, Hoydal MA, Wisloff U et al. Effects of preterm birth and fetal growth retardation on cardiovascular risk factors in young adulthood. *Early Human Development* 2009; 85(4):239-245.

Ferdynus C, Quantin C, Abrahamowicz M, Platt R, Burguet A, Sagot P et al. Can Birth Weight Standards Based on Healthy Populations Improve the Identification of Small-for-Gestational-Age Newborns at Risk of Adverse Neonatal Outcomes? *Pediatrics* 2009; 123(2):723-730.

Figueras F, Figueras J, Meler E, Eixarch E, Coll O, Gratacos E et al. Customised birthweight standards accurately predict perinatal morbidity. *Archives of Disease in Childhood Fetal & Neonatal Edition* 2007; 92(4):F277-F280.

Appendix 12. References list of included studies in systematic review of birth weight standards

Fitzhardinge PM, Steven EM, Fitzhardinge PM, Steven EM. The small-for-date infant. II. Neurological and intellectual sequelae. *Pediatrics* 1972; 50(1):50-57.

Gagliardo H-G. Visual function and fine motor control in small for gestational age infants. *Arq Neuropsiquiatr* 2004; 64(4):955-962.

Gardner DS, Hosking J, Metcalf BS, Jeffery AN, Voss LD, Wilkin TJ. Contribution of early weight gain to childhood overweight and metabolic health: a longitudinal study (EarlyBird 36). *Pediatrics* 2009; 123(1):e67-e73.

Glinianaia SV, Rankin J, Pearce MS, Parker L, Pless-Mulloli T. Stillbirth and infant mortality in singletons by cause of death, birthweight, gestational age and birthweight-for-gestation, Newcastle upon Tyne 1961-2000. *Paediatric and Perinatal Epidemiology* 2010; 24(4):331-342.

Gouyon B, Binquet C, Sagot P, Gouyon JB. Neonatal morbidity and mortality related to low birthweight in full-term neonates. *Pediatric Research* 2003; 53(4):2601.

Gray RH, Ferraz EM, Amorim MS, Demelo LF. Levels and Determinants of Early Neonatal-Mortality in Natal, Northeastern Brazil - Results of A Surveillance and Case-Control Study. *International Journal of Epidemiology* 1991; 20(2):467-473.

Haas JD, Balcazar H, Caulfield L, Haas JD, Balcazar H, Caulfield L. Variation in early neonatal mortality for different types of fetal growth retardation*. American Journal of Physical Anthropology* 1987; 73(4):467-473.

Hands B, Kendall G, Larkin D, Parker H. Perinatal Risk Factors for Mild Motor Disability. *International Journal of Disability Development and Education* 2009; 56(4):317-331.

Hemachandra AHK. The association between intrauterine growth restriction in the full-term infant and high blood pressure at age 7 years: Results from the Collaborative Perinatal Project. *International Journal of Epidemiology* 2006; 35(4): 871-877.

Hindmarsh PC, Bryan S, Geary MPP, Cole TJ. Effects of current size, postnatal growth, and birth size on blood pressure in early childhood. *Pediatrics* 2010;126 (6): e1507-e1513.

Indredavik MS, Vik T, Evensen KAI, Skranes J, Taraldsen G, Brubakk AM. Perinatal Risk and Psychiatric Outcome in Adolescents Born Preterm With Very Low Birth Weight or Term Small for Gestational Age*. Journal of Developmental and Behavioral Pediatrics* 2010; 31(4):286-294.

Jacobsson BA, Ahlin K, Francis A, Hagberg G, Hagberg H, Gardosi J. Cerebral palsy and restricted growth status at birth: Population-based case-control study. *BJOG: An International Journal of Obstetrics and Gynaecology* 2008; 115(10): 1250-1255.

Appendix 12. References list of included studies in systematic review of birth weight standards

Jelliffe-Pawlowski LLH, Hansen, R.L. Neurodevelopmental outcome at 8 months and 4 years among infants born full-term small-for-gestational-age. *Journal of Perinatology* 2004; 24(8): 505-514.

Johnson CC, Spitz MR. Nueroblastoma: case control analysis of birth characteristics. *Journal of the National Cancer Institute* 1985; 74:789-792.

Kajantie E, Osmond C, Barker DJP, Forsen T, Phillips DIW, Eriksson JG. Size at birth as a predictor of mortality in adulthood: a follow-up of 350 000 person-years. *International Journal of Epidemiology* 2005; 34(3):655-663.

Kindlund K, Thomsen SF, Stensballe LG, Skytthe A, Kyvik KO, Backer V et al. Birth weight and risk of asthma in 3-9-year-old twins: exploring the fetal origins hypothesis. *Thorax* 2010; 65(2):146-149.

Kotecha SJ, Watkins WJ, Heron J, Henderson J, Dunstan FD, Kotecha S. Spirometric Lung Function in School-Age Children Effect of Intrauterine Growth Retardation and Catch-up Growth. *American Journal of Respiratory and Critical Care Medicine* 2010; 181(9):969-974.

Kramer MSO, Olivier M, McLean FH, Willis DM, Usher RH. Impact of intrauterine growth retardation and body proportionality on fetal and neonatal outcome. *Pediatrics* 1990; 86(5):707-713.

Kuhle S, Allen AC, Veugelers PJ. Perinatal and childhood risk factors for overweight in a provincial sample of Canadian Grade 5 students. *International Journal of Pediatric Obesity* 2010; 5(1):88-96.

Lamb MM, Dabelea D, Yin X, Ogden LG, Klingensmith GJ, Rewers M et al. Early-life predictors of higher body mass index in healthy children. *Annals of Nutrition & Metabolism* 2010; 56(1):16-22.

Larroque B, Bertrais S, Czernichow P, Leger J. School difficulties in 20 year olds who were born small for gestational age at term in a regional cohort study. *Pediatrics* 2001; 108:111-115.

Levitt NS, Lambert EV, Woods D, Hales N, Andrew R, Seckl JR. Impaired glucose tolerance and elavated blood pressure in low birth rate weight, nonobese, young South African adults: Early programming of cortisol axis. *Journal of Clinical Endocrinology and Metabolism* 2000; 85(12): 4611-4618.

Li L, Law C, Power C. Body mass index throughout the life-course and blood pressure in mid-adult life: A birth cohort study. *Journal of Hypertension* 2007; 25(6): 1215-1223.

Libby G, Mcewan SR, Morris AD, Belch JJF. No difference in the association between birth weight and total cholesterol for males and females. A SHARP (Scottish

Appendix 12. References list of included studies in systematic review of birth weight standards

Heart and Arterial Disease Risk Prevention) study. *Vascular Medicine* 2008; 13(4):271-274.

Liew G, Wang JJ, Klein R, Duncan BB, Yeh HC, Brancati FL et al. Birth weight is not related to risk of diabetic retinopathy in type 2 diabetes: The atherosclerosis risk in communities study. *Current Eye Research* 2008; 33(2):193-198.

Lira PIC, Ashworth A, Morris SS. Low birth weight and mortality from diarrhoea and respiratory infection in NorthEast Brazil. *Journal of Pediatrics* 1996; 128: 497-504.

Low JA, Galbraith RS, Muir D, Killen H, Pater B, Karchmar J et al. Intrauterine growth retardation: a study of long-term morbidity*. American Journal of Obstetrics & Gynecology* 1982; 142(6 Pt 1):670-677.

Low JAH, Handley-Derry MH, Burke SO, Peters RD, Pater EA, Killen HL et al. Association of intrauterine fetal growth retardation and learning deficits at age 9 to 11 years. *American Journal of Obstetrics and Gynecology* 1992; 167(6): 1499-1505.

Lubchenco LO, Searls DT, Brazie JV. Neonatal mortality rate: relationship to birth weight and gestational age. *Journal of Pediatrics 1972*; 81(4):814-822.

Lurbe E, Carvajal E, Torro I, Aguilar F, Alvarez J, Redon J. Influence of concurrent obesity and low birth weight on blood pressure phenotype in youth. *Hypertension* 2009; 53(6):912-917.

Manji KP, Massawe AW, Mgone JM. Birthweight and neonatal outcome at the Muhimbili Medical Centre, Dar es Salaam, Tanzania. *East African Medical Journal* 1998; 75(7):382-387.

Mardones F, Marshall G, Viviani P, Villarroel L, Burkhalter BR, Tapia JL et al. Estimation of individual neonatal survival using birthweight and gestational age: a way to improve neonatal care. *Journal of Health, Population & Nutrition* 2008; 26(1):54-63.

Matijasevich A, Santos IS, Barros AJD, Menezes AMB, Albernaz EP, Barros FC et al. Perinatal mortality in three population-based cohorts from Southern Brazil: trends and differences. *Cadernos de Saude Publica* 2008; 24:S399-S408.

McIntire DD, Bloom SL, Casey BM, Leveno KJ. Birth weight in relation to morbidity and mortality among newborn infants. *New England Journal of Medicine* 1999; 340(16):1234-1238.

McKinney PA, Parslow R, Gurney KA, Law GR, Bodansky HJ, Williams R. Perinatal and neonatal determinants of childhood type 1 diabetes: A case-control study in Yorkshire, U.K. *Diabetes Care* 1999; 22(6): 928-932.

Appendix 12. References list of included studies in systematic review of birth weight standards

Meas T, Deghmoun S, Armoogum P, Alberti C, Levy-Marchal C. Consequences of being born small for gestational age on body composition: An 8-year follow-up study. *Journal of Clinical Endocrinology & Metabolism* 2008; 93(10):3804-3809.

Meas T, Deghmoun S, Alberti C, Carreira E, Armoogum P, Chevenne D et al. Independent effects of weight gain and fetal programming on metabolic complications in adults born small for gestational age. *Diabetologia* 2010; 53(5):907-913.

Melve KK, Skjaerven R, Melve KK, Skjaerven R. Birthweight and perinatal mortality: paradoxes, social class, and sibling dependencies. *International Journal of Epidemiology* 2003; 32(4):625-632.

Menezes AMB, Hallal PC, Horta BL, Araujo CLP, Vieira MD, Neutzling M et al. Size at birth and blood pressure in early adolescence: A prospective birth cohort study. *American Journal of Epidemiology* 2007; 165(6):611-616.

Mi JC, Cheng H, Zhao X-Y, Hou D-Q, Chen F-F, Zhang K-L. Developmental origin of metabolic syndrome: Interaction of thinness at birth and overweight during adult life in Chinese population. *Obesity Reviews* 2008; 9(Suppl. 1): 91-94.

Minakami H, Izumi A, Sato I. Gestational age-specific normal birth weight for Japanese twins - Risk of early neonatal death in small-for-gestational-age and large-for-gestational-age twins. *Journal of Reproductive Medicine* 1999; 44(7):625-629.

Minor VK, Divon MY. Fetal growth restriction at term: Myth or reality? *Obstetrics and* Gynecology 1998; 92(1):57-60.

Moser K, Macfarlane A, Chow YH, Hilder L, Dattani N, Moser K et al. Introducing new data on gestation-specific infant mortality among babies born in 2005 in England and Wales. *Health Statistics Quarterly* 2007;(35):13-27.

Nelson KB, Broman SH, Nelson KB, Broman SH. Perinatal risk factors in children with serious motor and mental handicaps. *Annals of Neurology* 1977; 2(5):371-377.

Nobili VM, Marcellini M, Marchesini G, Vanni E, Manco M, Villani A et al. Intrauterine growth retardation, insulin resistance, and nonalcoholic fatty liver disease in children. *Diabetes Care* 2007; 30(10):2638-2640.

North AF, Jr. Small-for-dates neonates. I. Maternal, gestational, and neonatal characteristics. *Pediatrics* 1966; 38(6):1013-1019.

O'Keeffe MJ, O'Callaghan M, Williams GM, Najman JM, Bor W, O'Keeffe MJ et al. Learning, cognitive, and attentional problems in adolescents born small for gestational age. *Pediatrics* 2003; 112(2):301-307.

Appendix 12. References list of included studies in systematic review of birth weight standards

Ott WJ. Small for gestational age fetus and neonatal outcome: reevaluation of the relationship. *American Journal of Perinatology* 1995; 12(6):396-400.

Owen CG, Whincup PH, Odoki K, Gilg JA, Cook DG. Birth weight and blood cholesterol level: a study in adolescents and systematic review. *Pediatrics* 2003; 111:1081-1089.

Padin-Rojas YY, Coll CGT, Gomex G, Bonet L, Escobar M, Varcarcel M. Intervencion Temprana en Infantes con Retraso en Creclmiento Intrauterino: Sus Efectos en el Desarrollo Neuropsicologico. *Boletin Assoc Med P Rico* 1991; 89(9):378-382.

Pandolfi C, Zugaro A, Lattanzio F, Necozione S, Barbonetti A, Colangeli MS. Low birth weight and later development of insulin resistance and biochemincal/ clinical features of polycystic ovarian syndrome. *Metabolism: Clinical and Experimental* 2008; 57:999-1004.

Pathai S, Cumberland PM, Rahi JS. Prevalence of and Early-Life Influences on Childhood Strabismus Findings From the Millennium Cohort Study. *Archives of Pediatrics & Adolescent Medicine* 2010; 164(3):250-257.

Paz I, Gale R, Laor A, Danon YL, Stevenson DK, Seidman DS. The cognitive outcome of full-term small for gestational age infants at late adolescence. *Obstetrics and Gynecology* 1995; 85:452-456.

Pearce MS, O'Sullivan JJ. Relationship between birth weight and blood pressure variability in children. *Journal of Human Hypertension* 2003; 17:677-680.

Peng Y, Huang B, Biro F, Feng L, Guo Z, Slap G. Outcome of low birth weight in China: a 16 year longitudinal study. *Acta Paediatrica* 2005; 94:843-849.

Plante LA. Small size at birth and later diabetic pregnancy. *Obstetrics and Gynecology* 1998; 92:781-784.

Pulver LS, Guest-Warnick G, Stoddard GJ, Byington CL, Young PC. Weight for gestational age affects the mortality of late preterm infants. *Pediatrics* 2009; 123(6):e1072-e1077.

Rahiala E, Tenhola S, Vanninen E. Ambulatory blood pressure in 12 year old children born small for gestational age. *Hypertension* 2002; 39:909-913.

Rich-Edwards J, Stampfer MJ, Manson J, Rosner B, Hankinson SE, Colditz GA et al. Birth weight and risk of cardiovascular disease in a cohort of women followed up since 1976. *BMJ* 1997; 315:doi:10.1136.

Salonen MK, Tenhola S, Laitinen T, Lvyra-Laitinen T, Romppanen J, Jaaskelainen J et al. Tracking serum lipid levels and the association of cholesterol conceentrations,

Appendix 12. References list of included studies in systematic review of birth weight standards

blood pressure and cigarette smoking with carotid artery intima-media thickness in young adults born small for gestational age. *Circulation Journal* 2010; 74: 2419-2425.

Sommerfelt KA, Andersson HW, Sonnander K, Ahlsten G, Ellertsen B, Markestad T et al. Behavior in term, small for gestational age preschoolers. *Early Human Development* 2001; 65(2):107-121.

Strauss RS. Adult functional outcome of those born small for gestational age: twenty six year follow up of the 1970 British birth cohort. *JAMA* 2000; 283:625-632.

Thompson C, Syddall H, Rodin I, Osmond C, Barker DJP. Birth weight and the risk of depressive disorder in late life. *British Journal of Psychiatry* 2001; 179:450-455.

Ullah A, Barman A, Haque J, Khanum M, Bari I. Birthweight and early neonatal health: Bangladesh perspective. *Paediatric and Perinatal Epidemiology* 2009; 23(6):542-547.

Uvebrant P, Hagberg G. Intrauterine growth in children with cerebral palsy*. Acta Paediatrica, International Journal of Paediatrics* 1992; 81(5): 407-12.

Vikse BE, Irgens LM, Leivestad T, Hallan S, Iversen BM. Low birth weight increases risk for end-stage renal disease. *Journal of the American Society of Nephrology* 2008; 19(1):151-157.

Walther FJ, Ramaekers LH. Language development at the age of 3 years of infants malnourished in utero. *Neuropediatrics* 1982; 13(2):77-81.

Walther FJ, Ramaekers LH. The ponderal index as a measure of the nutritional status at birth and its relation to some aspects of neonatal morbidity. *Journal of Perinatal Medicine* 1982; 10(1):42-47.

Wei JN, Li HY, Chang CH, Sung FC, Li CY, Lin CC et al. Birth weight and type 1 diabetes among schoolchildren in Taiwan--A population-based case-controlled study. *Diabetes Research & Clinical Practice* 2006; 74(3):309-315.

Wennergren M. Perinatal risk factors. With special reference to intrauterine growth retardation and neonatal respiratory adaptation*. Acta Obstetricia et Gynecologica Scandinavica - Supplement* 1986; 135:1-51.

Zhang X, Cnattingius S, Platt RW, Joseph KS, Kramer MS. Are babies born to short, primiparous, or thin mothers "normally" or "abnormally" small? *Journal of Pediatrics* 2007; 150(6):603-607.

Appendix 13. Search strategy for systematic review of Apgar score and adverse outcomes

1. Apgar score
2. apgar score.mp
3. apgar.mp
4. Resuscitation
5. Neonatal resuscitation
6. neonat*.mp.
7. infant*.mp.
8. newborn.mp.
9. 1 or 2 or 3 or 4 or 5
10. 6 or 7 or 8
11. 9 and 10
12. Diabetes Mellitus/
13. diabetes mellitus.mp.
14. Hypertension/ or hypertension.mp.
15. cardiovascular disease.mp. or Cardiovascular Diseases/
16. Metabolic Syndrome X/ or metabolic syndrome.mp.
17. cerebral palsy.mp. or Cerebral Palsy/
18. Developmental Disabilities/
19. developmental delay.mp.
20. learning difficulties.mp.
21. Learning Disorders/
22. neonatal morbidity.mp.
23. Infant Mortality/
24. neonatal mortality.mp.
25. infant mortality.mp.
26. 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25
27. 11 and 26
28. limit 27 to humans

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Author, Year and Country	Study design	Population (total number)	Gestational age and method of defining	Apgar score timing and threshold	Outcome Measure
Abrams et al 2007 USA	Retrospective cohort	(598) All infants with hydrops fetalis in neonatal network, any cause, born Jan 1996-March 2005 Congenital anomalies not excluded	unreported	1 min ≤ 3 5 min ≤ 3	Neonatal death (exact range unclear, death prior to discharge from hospital)
Adamson et al 1995 Australia	Case control study	(176) All singleton term infants in metropolitan area admitted to 2 study hospitals in 1992 (8 month period) within 1st week of life with diagnosis of neonatal encephalopathy, plus one matched control on sex, hospital of delivery, time of birth, day of the week, maternal health insurance status.	≥ 37 weeks Method unreported	1 min ≤ 3 1 min ≤ 7	Neonatal encephalopathy (1^{st} 7 days of life) (at least one of the following: seizures of any type/ duration, absent responsiveness to stimuli (stupor/ coma); altered responsiveness to stimuli for > 24 hrs, difficulty with control of respiration (presumed brain stem origin) inc cyanotic attacks after 2 days of age and recurrent apnoea any age, poor suck)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Aggarwal et al 2005 India	Case control study	(50) Consecutive survivors at ≥ 34 weeks, 5 min Apgar ≤ 6 or needed resuscitation for ≥ 5 mins, and 25 matched controls without asphyxia (Apgar ≥ 7) for gestation and weight, born at a single tertiary centre Congenital anomalies and neonates exposed to nephrotoxic drugs excluded	≥ 34 weeks Method unreported	5 min ≤ 3 5 min ≤ 6	Acute renal failure (neonatal) (serum creatinine >1.5mg/dl)
Ajayi and Nzeh 2003 Nigeria	Prospective cohort	(43) Babies with birth weight <1500g at delivery who survived at least 72 hours born at a secondary hospital 1992-1994	Unreported	1 min ≤ 3 5 min ≤ 3	Intraventricular haemorrhage/ periventricular leukomalacia (neonatal) (Papile Grade 1-4 on cranial USS)
Ambalavanan and Carlo 2001 USA	Retrospective cohort	(811) Neonates with birth weight <1000g neonates, excluding those who died in the delivery room, admitted to a tertiary care centre, 1990-1996. Congenital anomalies excluded	Unreported	5 min, all thresholds	Neonatal death (before discharge, exact age unspecified)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Anyaegbunam et al 1991 USA	Prospective cohort	(270) Live births with umbilical cord pH taken and delivery on dates when investigator attending, all singletons with cephalic presentation	>36 weeks gestation Method unreported	5 min < 7	Neonatal sepsis (definition unreported)
Apgar and James 1962 USA	Cohort, unclear if prospective/ retrospective	(27715) All live births > 500g at a single centre, 1952-1959	Unreported	1 min, all thresholds	Neonatal death (28 days)
Atkinson 1983 USA	Retrospective cohort	(239255) All live births in North Carolina region, 1978-1980. Data presented separately for birth weight <1501g	Unreported	5 min ≤ 3 5 min ≤ 6	Neonatal death (28 days)
Baenziger et al 1999 Switzerland	Prospective cohort	(12) High risk population, all had CTG abnormalities (bradycardia to <80bpm, reduced variability, late decelerations)or meconium or Apgar <6, born at single tertiary centre 1993-1995. Congenital anomalies excluded	34+5-42+1 weeks gestation Method unreported	1 min ≤ 3	Griffiths developmental quotient (≤ 80 abnormal) (9-15 months)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Bauder, von Siebenthal and Bucher 2000 Switzerland	Case control study	(72) 40 infants with PVL were registered to Swiss Pediatric surveillance unit, over 3 years. They were compared to a control group matched for gestational age. 35 were preterm infants 1995-1997	Unreported	5 min <7	Periventricular leukomalacia (at least 2 cysts with diameter 2mm in the periventricular region)(neonatal)
Beeby et al 1994 Australia	Retrospective cohort	(623) Infants <32 weeks gestation born at a single centre. Year of birth 1985-1990 Congenital anomalies excluded	<32 weeks gestation Unreported	1 min <4	1.Neonatal death (exact age unspecified) 2.Intraventricular haemorrhage (Gd 3-4 Papile, cranial USS or post-mortem)(neonatal, up to 14 days) 3. Cerebral palsy (definition unreported)(age 1 year)
Behnke et al 1987 USA	Prospective cohort	(748) live born infants with birth weight 500-1800g, born at a single centre 1974-1980.	Range unreported LMP/ Dubowitz method	1 min ≤ 3 1 min ≤ 3 5min ≤ 6 5 min ≤ 6	Neonatal death (28 days)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Behnke et al 1989 USA	Prospective cohort	(256) Same study population as above, 20% sample selected for long term follow up	Range unreported LMP/ Dubowitz method	1 min ≤ 3 1 min ≤ 3 5min ≤ 6 5 min ≤ 6	1.Bayley mental development index (abnormal <85) (12 months of age) 2.Bayley psychomotor development index (abnormal <85) (12 months of age)
Bennett, Robinson and Sells 1983 USA	Retrospective cohort	(8) All newborns of birth weight <800g admitted to tertiary neonatal care unit, 1977-1980, who survived to discharge home and were at least 3 years of age at time of report No exclusions reported	24-28 weeks gestation LMP/ Dubowitz method	1 min ≤3	1.Bayley mental development index (abnormal <85) (6 months-2years of age) 2.Stanford-Binet IQ score (abnormal <100) (age 3 years)
Berger et al 1997 Germany	Prospective cohort	(5280) All live born infants at a single tertiary centre from 1984 to 1988. No exclusions reported	24-43 weeks gestation Method unreported	1 min ≤ 4 1 min ≤ 7 5 min ≤ 4 5 min ≤ 7 10 min ≤ 4 10 min ≤ 7	Periventricular/ intraventricular haemorrhage (Grade 1-3 Papile, cranial USS) (neonatal day5-8)
Brandalise et al 1976 Brazil	Prospective cohort	(93) newborns with different Apgar scores, unclear how selected. Birth weight >2500g. Year of birth unreported. No exclusions reported.	≥ 37 weeks gestation Method unreported	5 min < 7	Neonatal death (exact age unreported)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Buchmayer et al 2009 Sweden	Case control study	(7160) Cases were infants with autistic disorders who had had hospital treatment, 5 randomly selected controls matched to gender, birth year and birth hospital. Year of birth 1987-2002. Congenital malformations NOT excluded.	Unreported	5 min ≤6	Autistic disorders (ICD-9/10 classification) (up to age 10 years)
Camp et al 1998 USA	Prospective cohort	(33934) Collaborative Perinatal project, mothers enrolled antenatally in one of 12 urban medical centres in the USA, 1959-1965. Infants had birth weight > 2000g Congenital anomalies, syndromes or infants who had neonatal seizures excluded	Unreported	1 min <2	Weschler intelligence scale for children (IQ <70)(age 7 years)
Casey, McIntire and Leveno 2001 USA	Prospective cohort	(12899) Live born singleton infants at a single tertiary centre 1988-1998, excluded if no cord pH available, no other exclusions reported.	>26 weeks gestation Method unreported	5 min ≤3 5 min ≤ 6	Neonatal death (28 days)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Chandola et al 1992 UK	Retrospective cohort	(24672) Index cases from referral register of children's clinic, who presented with hyperactivity, were free from mental handicap, psychoses and gross neurological impairment, were not referred for any other problem. 129 could be linked on Cardiff Births Survey register, compared with other births in South Glamorgan region born 1980-1984. No exclusions reported	Unreported	1 min ≤6 5 min ≤8	Referral for hyperactivity symptoms (predominance of restless, inattentive and chaotic behaviour) (age 3-6 years)
Colburn and Salzman 1960 USA	Cohort, unclear if prospective/ retrospective	(1597) Consecutive deliveries of infants with birth weight >1000g at a single centre. Year of birth 1959 to 1960 No exclusions reported	Unreported	1 min ≤ 3 1 min ≤ 6	Neonatal death (exact age unspecified)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Dalens et al 1981 France	Prospective cohort	(57) neonates admitted to NICU at a single centre, without CSF infection, necessity for several examinations of CSF, precise knowledge of gestation and obstetric states, consenting to long term follow up Year of birth unreported No exclusions reported	37-42 weeks Method unreported	1 min <3 1 min ≤ 6	Neurological sequelae (threshold unreported) (age 12 months)
De Almeida et al 2008 Brazil	Prospective cohort	(579) All live births of infants with birth weight 400-1500g and gestation 22-33 weeks, at 8 tertiary hospitals. Year of birth 2004-2005 Congenital anomalies excluded	22-33 weeks Obstetric estimate/ physical examination of infant	5 min <7	Neonatal death (first 7 days)
Den Ouden et al 1990 Netherlands	Prospective cohort	(1192) All live born infants in Netherlands with gestational age < 32 weeks and/ or birthweight <1500g. Year of birth 1983 Congenital anomalies excluded	<32 weeks	5 min ≤ 6	Neurological dysfunction (increased/ decreased excitablility, increased/ decreased mobility, hyper/hypotonia, asymmetry)(neonatal)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Deorari, Paul and Singh 1989 India	Prospective cohort	(103) 36 infants with Apgar ≤ 3, 32 infants with Apgar 4-6, and 35 matched for weight and gestation who had 1 min apgar >6. Born at single centre, year of birth unreported. No exclusions reported	Unreported	1 min ≤ 3	1. Seizures (threshold unreported) (up to 72 hours of age) 2.Developmental delay (DQ <70) (age 12 months)
Dijxhoorn et al 1986 Netherlands	Prospective cohort	(803) Infants part of Groningen Perinatal project, delivered vaginally 1975-1978, average weight for gestational age, at term gestation. Year of birth 1975-1978 No exclusions reported	≥ 37 weeks Method unreported	1 min ≤ 3 1 min ≤ 6	Neurological abnormality (hyper/hypokinesia, hyper/hypotonia, hemisyndrome, apathy, hyperexcitability) (neonatal day 4-5)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Drage et al 1966 USA	Prospective cohort	(18038) Collaborative Study of Cerebral Palsy, Mental Retardation and other Neurological and Sensory Disorders of Infancy and Childhood. 13 collaborating institutions, year of birth unreported. Exclusions: malformations of CNS, Down's syndrome, hypothyroidism, other syndromes associated with mental retardation	Unreported	5 min ≤ 6	1.Abnormal gross motor skills (walking along a straight line, hopping, ball catching)(age 4 years) 2. Abnormal behaviour profile(5 point scale: emotional reactivity, degree of irritability, degree of cooperation, degree of dependence, duration of attention span, goal orientation, response to direction, level of activity, nature of activity, nature of communication) (age 4 years)
Drage et al 1964 USA		As above	Unreported	1 min ≤ 3 1 min ≤ 6	Neonatal death (28 days)
Drage et al 1966 USA		As above	Unreported	5 min ≤ 3 5 min ≤ 6	Neurological abnormality (definition unreported) (age 12 months)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Echandia and Ruiz 2006 Colombia	Retrospective cohort	(287) Low socioeconomic stratum Infants with one of the following risk factors: respiratory problems requiring mechanical ventilation, Apgar <7 at 5 minutes, 3 apnoeic episodes, neurological abnormalities such as convulsions, paralysis, hyper/ hypotonia, altered consciousness, bacterial meningitis. Single level 1 hospital, year of birth 1989 to 1997. Exclusions: congenital anomalies, intrauterine infection, prenatal neurological compromise	27-43 weeks gestation Method unreported	1 min ≤ 3 1 min ≤ 6 5 min ≤ 3 5 min ≤ 6	1.Seizures (definition unreported) (neonatal) 2. Neuromotor abnormalities (definition unreported) (age 12 months)
Ehrenstein et al 2006 Denmark	Retrospective cohort	(131696) All singleton live births in 1978-2001 in North Jutland County, Denmark. Congenital anomalies excluded	No gestational age restrictions Method unreported	5 min <7	Hospitalisation with epilepsy (Hospital discharge database ICD-10 codes) (childhood, median age 12 years)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Ekure, Iroha and Egri-Okwaji 2011 Nigeria	Prospective cohort	(560)Consecutive deliveries at a single university teaching hospital (stillbirth reported in paper but only live births in analysis). 86.1% term, 13.6% preterm. Year of birth 2002 (6 month period) No exclusions reported	No gestational age restrictions Method unreported	1 min ≤ 3 1 min ≤ 6 5 min ≤ 3 5 min ≤ 6	Neonatal death (7days)
Evans et al 2007 Australia	Retrospective cohort	(5713) All infants of 24-32 weeks gestation, data collected through Australia and New Zealand neonatal network prospective audit. Year of birth 1998-2001 Congenital anomalies and hydrops fetalis excluded	24-32 weeks gestation Early obstetric USS/ LMP	1 min <4	Neonatal death (before discharge from hospital)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Graham, Holcroft and Blakemore 2002 USA	Case control study	(36) All cases of neonatal seizures according to hospital records, matched 2:1 by birth weight, gestational age and route of delivery to neonates born during the same period at a single tertiary centre. Year of birth 1988-1999 Congenital anomalies excluded	24-40 weeks gestation Method unreported	1 min <7 5min <7	Seizures (definition unreported) (neonatal)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Grether et al 1996 USA	Case control study	(114) Cases were singleton births during study period in one of four San Francisco bay counties, birth weight <1500g at birth, survival to age 3 years, and moderate or severe CP. Controls randomly selected, without matching, from infants with birth weight <1500g born in same counties and survived to age 3 years Year of birth 1983-1985 Excluded if CP from cause after 28 days, or transient abnormalities. Congenital anomalies/ infections not excluded from analysis	Gestation unreported LMP/ USS before 19 weeks gestation	5 min ≤ 5	Cerebral palsy (chronic disability of CNS origin characterised by aberrant control of movement or posture, appearing early in life and not the result of progressive disease)(age 3 years)
Heller et al 2003 Germany	Retrospective cohort	(512496) Data obtained from the perinatal birth register of Hesse, all live births within study period. Year of birth 1990-1999 Congenital anomalies excluded	No restrictions on gestational age Method unreported	1 min ≤ 3 1 min ≤ 7 5 min ≤ 3 5 min ≤ 7 10 min ≤3 10 min ≤ 7	Neonatal death (7 days)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Heuchan et al 2002 Australia	Retrospective cohort	(5637) Data from Australia and New Zealand neonatal network, live births during study period 24-30 weeks gestation with USS or post-mortem data available Year of birth 1995-1997 Excluded if data unavailable	24-30 weeks gestation Method unreported	1 min <4	Intraventricular haemorrhage (cranial USS/ post-mortem Papile grade 3-4)(neonatal)
Holst et al 1989 Denmark	Retrospective cohort	(4038) All surviving singletons born in Frederiksborg county during 1978, records from birth and routine check-up at 4 years old. Congenital anomalies not excluded	Unreported	1 min ≤6 10 min <10	Handicap (cerebral palsy, mental retardation (mild/severe), epilepsy, severe defects of vision and hearing. GP records, exact definitions unreported) (age 4 years)
Iijima et al 2009 Japan	Retrospective cohort	(113) Live born infants 22-24 weeks gestation who survived delivery room resuscitation and were treated with expectation of survival at a single tertiary centre. Year of birth 1991-2006 Congenital anomalies and lethal chromosomal anomalies excluded	Early USS (95%) and LMP	5 min ≤ 3	1.Neonatal death (before discharge from hospital) 2. Neurological disability (cerebral palsy, developmental delay, deafness and blindness. Exact definitions unreported, from hospital records) (age 2 years)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Ikonen 1967 Finland	Retrospective cohort	(11855) Live born infants in single centre during study period. Year of birth 1963-1965 No exclusions reported	Unreported	1 min ≤3 1 min ≤ 6	Neonatal death (exact age unreported)
Ikonen 1973 Finland	Retrospective cohort	(1006)Series selected from among infants with birth weight >2500g: all infants with apgar 0-3, all infants with apgar 4-6, and a random selection of infants with scores 7-10 born at a single centre. Year of birth 1964-1968 No exclusions reported	Unreported	1 min ≤3 1 min ≤ 6	1.Cerebral disturbances (intracranial haemorrhage (clinical/autopsy), cerebral lesion or clinical suspicion, cerebral irritation, cerebral depression, symptomatic hypoglycaemia, hypocalcaemia. Exact thresholds unreported) (neonatal) 2. Respiratory disturbances (including aspiration syndrome (resp sx with radiological findings of streaky infiltration in perihilar region), hyaline membrane disease (Silverman's score ≥ 4/ radiographic findings), pneumonia, respiratory distress (resp rate >70/min for at least 6 hrs with Silverman's score <4)) (neonatal)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Ikonen 1974 Finland	Retrospective cohort	(218) Series selected from among live born infants BW 1000-1990g all infants with 1 min Apgar 0-3, all infants with apgar 4-6, random sample of infants with score 7-10. Year of birth 1964-1968. No exclusions reported	Unreported	1 min ≤3 1 min ≤ 6	As above
Indredavik et al 2010 Norway	Prospective cohort	(192) 3 groups: preterm, birth weight ≤ 1500g, term small for gestational age(birth weight <10th centile), and term normal weight controls. All VLBW admitted to NICU at Trondheim university hospital 1986- 1988, term SGA and control infants born to mothers living in city of Trondheim, enrolled into study 10% random sample controls, and all SGA children included Congenital anomalies excluded	Preterm 24-41 weeks Term ≥ 37 weeks gestation Method unreported	1 min ≤3 1 min ≤ 6 5 min ≤3 5 min ≤ 6	Psychiatric diagnoses (Schedule for affective disorders and schizophrenia for school age children (KSADS), DSM-IV criteria) (age 14 years)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Ishikawa et al 1995 Japan	Prospective cohort	(164) Cohort of infants with birth weight 490-1500g born at a single tertiary centre. Year of birth 1977-1982 No exclusions reported.	Unreported	5 min <4	Epilepsy (chronic condition characterised by the repeated occurrence of seizures such as those with EEG epileptic seizure discharges) (age 6 years)
Issel et al 1976 Germany	Retrospective cohort	(6483) Regional cohort of infants, birth weight 1010-4500g. Year of birth unreported, no exclusions reported	Unreported	1 min ≤ 3 1 min ≤ 7	Neonatal death (exact timing unreported)
Jacobsson et al 2002 Sweden	Case control study	(435) All preterm children with spastic CP if born and lived in study region(western Sweden) for 4 years and lacked obvious postnatal cause of CP. Matched with 2 controls, closest births before and after case birth matched for gestational age, gender, multiple gestation and delivery ward. Year of birth 1983-1990 No exclusions reported	<37 weeks gestation 97% USS performed 16-19 weeks gestation	1 min <7 5 min <7 10 min <7	Cerebral palsy (group of non- progressive motor impairment syndromes, secondary to lesions or abnormalities of the brain arising in the early stages of development)(age 4 years and older)
Jennett et al 1981 USA	Retrospective cohort	(10124) All live births at single hospital during study period. Year of birth 1977-1980 No exclusions reported	Unreported Obstetric estimate/ Ballard examination of the newborn	1 min and 5 min, all thresholds	Neonatal death (exact age unreported)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Johnson et al 1999 USA	Prospective cohort	(28127) Infants with birth weight ≥ 2500g and jaundice, not treated with exchange transfusion or phototherapy. Year of birth 1959-1966. No exclusions reported	Unreported	5 min ≤3 5 min ≤ 6	Neurological disability (suspected/ confirmed, exact definition unreported) (age 7 years)
Kato et al 1996 Japan	Retrospective cohort	(228) Cohort of singleton infants with birth weight <1500g born at 2 centres 1984-1993. Major anomalies excluded	Unreported	1 min ≤ 4 5 min ≤ 4	1.Neonatal mortality (7 days) 2. Cerebral palsy/ mental retardation (definition unspecified) (>12 months old, exact age unspecified)
Krebs, Langhoff-Roos and Thorngren-Jerneck 2001 Denmark	Case control study	(274) All singleton breech presentation at term, identified from national registry data. 115 infants with Apgar score ≤6 at 5 min and controls (subsequent 2 deliveries from same hospital with Apgar >6 at 5 min) selected Year of birth 1982-1992 Congenital anomalies excluded	>37 weeks gestation Method unreported	5 min ≤ 6	1.Cerebral palsy 2. Speech and language problems (definitions unreported, questionnaire to parents) (age 4-15 years)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Lee, Subeh and Gould 2010 USA	Retrospective cohort	(690933) All singleton live births of preterm infants from US birth cohort dataset (National Center for health statistics). Year of birth: 2001-2002 Exclusions: congenital anomalies, multiple pregnancy, extremes of birth weight (<1st or >99th percentile)	24-36 weeks gestation LMP	5 min ≤3 5 min ≤ 6	Neonatal death (28 days)
Lie, Groholt and Eskild 2010 Norway	Retrospective cohort	(543064) All singletons, born during study period, who survived the first year of life. Norwegian birth registry data. Year of birth 1986-1995 Congenital anomalies excluded	Unreported	5 min ≤4 5 min ≤6	Cerebral palsy (ICD-9 classification) (up to age 5 years)
Luthy et al 1987 Canada	Prospective cohort	(246) Singleton infants, cephalic presentation, with birth weight 600-1750g born at single tertiary centre during study period. Year of birth 1981-1985 Exclusions: congenital anomalies, antenatal haemorrhage, infants delivered before the onset of labour.	26-32 weeks Method unreported	1 min ≤3	1.Neonatal mortality (exact age unreported) 2.Intracranial haemorrhage (cranial USS, Papile grade 3-4) (neonatal, 76-92 hours) 3. Cerebral palsy (definition unreported) (18 months of age)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Mahon et al 2007 Denmark	Retrospective cohort	(335268) All singleton live births in North Jutland, Aarhus and Viborg counties of Denmark, obtained from national registry data. Year of birth 1980-2001 Congenital anomalies NOT excluded. Excluded implausible gestational age/ birth weight combination i.e. <28 weeks and >2500g	25-45 weeks gestation Method unreported	5 min ≤ 6	Hospital admission with pneumococcal disease (ICD 8/10 codes pneumoccocal meningitis, septicaemia, pneumococcal pneumonia) (age 0-144 months)
Minchom et al 1987 UK	Retrospective cohort	(41144) Cases with seizures within 48 hours of birth, and other infants born during the study period who had not had seizures. Infants identified by data from the Cardiff Births survey. Year of birth: 1970-1979 Congenital anomalies NOT excluded	≥ 37 weeks Method unreported	1 min ≤ 3	Seizures (convulsions, fits, seizures, clonic movements or jerky movements recorded in case notes by experienced observers) (neonatal, first 48 hours)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Misra et al 1994 India	Retrospective cohort	(154) Cases were term infants with Apgar score ≤ 6, controls had Apgar ≥ 7, born at a single centre. Year of birth unreported Exclusions: history of maternal narcotic intake, respiratory distress, central cyanosis, jaundice, abdominal distension and congenital malformations	37 -41 weeks Method unreported	5 min ≤3 5 min <7	1.Neonatal death (exact age unreported) 2. Denver developmental screening test (abnormal, threshold unreported) (age 3 months and age 11 months)
Moro et al 2007 Spain	Retrospective cohort	(8741) Infants with birth weight <1500g from Spanish Society of Neonatology (SEN) 1500 database Year of birth: 2002-2005 No exclusions reported	Gestation range unreported LMP	1 min <4 1 min ≤ 6 5 min <4 5 min ≤ 6	1.Mortality (prior to discharge, up to 60 days)
Moster, Lie and Markestad 2001 Norway	Retrospective cohort	(235165) All infants born during study period with birth weight > 2500g, or if birth weight data missing gestational age >37 weeks from Norwegian birth registry. Year of birth: 1983-1987 Congenital anomalies excluded except congenital dislocation of the hip	Unreported	5 min ≤ 3 5 min ≤ 6	1.Neonatal death (exact age unreported) 2. Infant death (first year of life) 3. Cerebral palsy (definition unreported) (8-12 years old)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Moster et al 2002 Norway	Retrospective cohort	(727) All survivors with children with birth weight >2500g and Apgar 0-3, a random sample of 400 with score 4-6 and 404 score 7-10, identified from Norwegian birth registry Year of birth: 1983-1987 Congenital anomalies excluded except congenital dislocation of the hip, major neurological impairment also excluded.	Unreported	5 min ≤ 3 5 min ≤ 6	1.Seizures (definition unreported, questionnaire to parent when child 8-13 years old)(neonatal) 2. Minor motor impairment (definition unreported, questionnaire to parents)(8-13 years old) 3. Epilepsy (definition unreported, questionnaire to parents) (8-13 years old)
Moura et al 2010 Brazil	Prospective cohort	(3845) Live births in Pelotas birth cohort, born in 2004. No exclusions reported	Gestation range unreported Dubowitz method	5 min <7	Suspected developmental delay (Battelle Screening developmental inventory (BSDI), cut off -1SD in score table for reference population) (age 12 months)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Murphy et al 1995 UK	Case control study	(293) Cases all singleton children with cerebral palsy identified from regional register of early childhood impairments, controls surviving singleton preterm infants randomly selected from hospital admission registers and Oxford record linkage study. Year of birth: 1984-1990 No other exclusions reported	<32 weeks LMP and USS prior to 20 weeks gestation	5 min ≤ 3	Cerebral palsy (permanent disorder of movement and posture) (age 3-5 years)
Myers, Paton and Fisher 1987 USA	Retrospective cohort	(226) All live born infants 22-32 weeks gestation or 500-1499g included, born at a single centre. Year of birth 1983-1984 Congenital anomalies incompatible with life excluded	22-32 weeks gestation LMP, size at first obstetric examination, quickening	1 min ≤ 3	Death (neonatal/postneonatal)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Nelson and Ellenberg 1981 USA	Prospective cohort	(47869) Infants born to women participating in the National Collaborative Perinatal project, received antenatal care and delivered at one of 12 teaching hospitals, singleton children with known Apgar scores included. Year of birth: 1959-1966 Exclusions: cerebral palsy resulting from trauma, infection or vascular accident after 1st month of life or gross CNS malformations	Unreported	1 min ≤ 3 1 min ≤ 6 5 min ≤ 3 5 min ≤ 6 10 min ≤ 3 10 min ≤ 6	1. Cerebral palsy (chronic disability characterised by aberrant control of movement or posture appearing early in life, not the result of recognised progressive disease. Significant handicap in independent functioning) (age 1-7 years) 2. Infant mortality (first year of life)
Oain et al 1988 Norway	Retrospective cohort	(580) All singleton breech deliveries, birth weight >500g, at a single tertiary centre. Year of birth 1972-1979 Congenital anomalies NOT excluded	Unreported	1 min ≤ 3 1 min ≤ 6 5 min ≤ 6	Neonatal death (exact age unreported)
Obwegeser, Bohm and Gruber 1993 Germany	Retrospective cohort	(7848) All live births within study period, at single university hospital, with Apgar score and cord pH available. Year of birth: 1981-1990. No exclusions reported	Unreported	5 min ≤7	Neonatal death (exact age unreported)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Paludetto et al 1981 Italy	Prospective cohort	(50) Neonates admitted to NICU at a single tertiary centre for a variety of reasons, 50 of 138 admitted during the study period followed up. Year of birth 1975-1976 No exclusions reported	≥ 37 weeks gestation Method unreported	1 min ≤ 3 5 min ≤ 7	Neurodevelopmental delay (any including tetraplegia, psychomotor delay, hemiparesis, West's syndrome. Vojta's diagnostic scheme, Brunet Lezine's scale; threshold unreported) (age 9-24 months)
Patterson et al 1994 UK	Case control study	(1626) Cases with IDDM born in Scotland during the study period identified from computerised hospital discharge records, with an additional 21 cases from a database maintained by Scottish clinicians. Record linkage to match to maternal discharge records to obtain perinatal data. 5 control subjects randomly selected from deliveries of same sex on same date at hospitals in the same health board. All singletons. Year of birth: 1975-1976 No other exclusions reported	no restrictions, 8 % of cases/ controls <38 weeks gestation Method unreported	5 min ≤ 7	Insulin dependent diabetes mellitus (diagnostic criteria unreported) (age <15 years)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Paul, Koh and Monfared 1979 USA	Retrospective cohort	(201) All live births 1001-1500g at single centre. Year of birth 1975-1977 Excluded congenital anomalies incompatible with life, congenital infections	Unreported	5 min <7	Neonatal death (exact age unspecified)
Perlman and Risser 1996 USA	Prospective cohort	(96) Neonates born at a single centre and admitted to NICU who were at high risk for asphyxia: meconium, CTG abnormalities, abruption or 5 min Apgar 5 or less, pH <7.0 or BE >14, requiring ventilation or chest compressions. Year of birth: 1993 No exclusions reported	>37 weeks gestation Method unreported	5 min ≤5	Seizures (definition unreported) (day 1 of life)
Robertson and Harrild 2010 UK	Case control study	(1442) Cases of type 1 diabetes under 15 years old from the SSG for the care of Diabetes in the young database, linked to data on the Aberdeen maternity neonatal databank, matched to 3 controls according to year of birth. All singletons Year of birth: 1972-2002 No other exclusions reported	Unreported	1 min ≤ 7 5 min ≤ 7	Insulin dependent diabetes mellitus (diagnostic criteria unreported) (age 0-15 years)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Ruth and Raivio 1988 Finland	Prospective cohort	(925) All live born infants at a single centre over a 2 month period. Year of birth: 1984 Congenital anomalies not excluded initially, but abnormal outcome related to cause other than asphyxia excluded from the analysis	Unreported	5 min ≤ 7	Death or abnormal development (definite abnormality including cerebral palsy or noticeable delay in development, slight abnormality if transient delay in muscle tone, slightly abnormal development or abnormal pattern of motor function development) (12 months of age)
Sanders and Slade 2010 Australia	Retrospective cohort	(882) A sample of 5 year old children completed questionnaire and transcribed information from birth records, sampling frame electronic database of School dental service in South Australia. All singletons Year of birth: 1998 No other exclusions reported	37-41 weeks LMP/USS	5 min ≤ 8	Dental caries (one or more decayed, missing or filled teeth) (age 5 years)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Schmidt, Schuz and Lahteenmaki 2010 Denmark, Finland, Sweden and Norway	Case control study	(14518) Cases children diagnosed with a primary CNS tumour. Identified from national cancer registries. Matched by DOB, sex, country to 5 control children aged 0-14 years. Year of birth: 1971-2006. No exclusions reported	Gestation range unreported LMP/ early scan after period in study where introduced into general practice	5 min ≤ 7	CNS tumour (International Classfication of Childhood cancer version 3) (age 0-14 years)
Seidman et al 1991 Israel	Retrospective cohort	(1942) All subjects born at single tertiary centre between 1970 and 1971 and subsequently drafted into the army 17 years later. All singletons No other exclusions reported	Unreported	1 min ≤ 7 5 min ≤ 7	Intelligence test (equivalent of Weschler scale, threshold <85) (age 17 years)
Serenius et al 2004 Sweden	Retrospective cohort	(218) All live born infants of 23-25 weeks gestation in study period, born at 2 tertiary centres. Year of birth 1992-1998 Congenital anomalies NOT excluded	23-25 weeks gestation <18 week scan 94%, LMP 6%	1 min ≤ 3 5 min ≤ 3	Death (prior to discharge from neonatal unit, exact age unspecified)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Shankaran et al 2002 USA	Retrospective cohort	(5986) National institute of child health and development (NICHD) research network registry (12 centres). All live births 501-1000g. Year of birth: 1993-1997 No exclusions reported	Gestation range unreported LMP, obstetric examination and USS	1 min ≤ 3 5 min ≤ 3	Death (up to 120 days of age)
Soleimani et al 2010 Iran	Case control study	(3577) Cases were children with documented cerebral palsy, referred to rehabilitation centre, controls 3465 children without CP who had attended centre for well-being check-up. Year of birth: 2001-2007 Exclusions: no overt congenital anomalies, chromosomal, metabolic and neurodegenerative disorders, congenital infections	Gestation range unreported LMP	5 min <5	Cerebral palsy (non-progressive motor dysfunction, examination findings of increased tone (spasticity, rigidity, dystonia, or choreoathetosis) (age1-6 years)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Stelmach, Pisarev and Talvik 2005 Estonia	Case control study	(421) Cases children with cerebral palsy of any severity, ascertained from population based prevalence study (children aged 1-15) and controls selected from general population register. Matched by sex, year and month of birth and place of residence at the time of birth. Year of birth: 1985-1999 Exclusions: postnatal aetiology of CP, syndromes and CNS malformations	27-42 weeks gestation Method unreported	1 min ≤ 4 1 min ≤ 7 5 min ≤ 7	Cerebral palsy (permanent disorder, movement and posture, non-progressive lesion, presence of lesion in developing/ immature brain) (age 1-15 years
Tejani and Verma 1989 USA	Retrospective cohort	(392) All infants with birth weight <2000g and umbilical cord pH available during study period. Year of birth: 1981-1986 Congenital anomalies excluded	Gestation range unreported Obstetric estimate and Dubowitz method	1 min ≤ 6 5 min ≤ 6	1.Neonatal death (exact age unreported) 2. Peri/intraventricular haemorrhage (Papile grade 1-4 on cranial USS) (1st 24 hours of life) 3. Respiratory distress syndrome (radiologic evidence of reticulogranular pattern/ air bronchograms) (neonatal)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Thorngren-Jerneck and Herbst 2001 Sweden	Retrospective cohort	(1028705) Infants born at term during study period, data from Swedish Medical Birth registry Year of birth: 1988-1997 Congenital anomalies excluded	≥ 37 weeks gestation USS/ LMP	5 min <7	1.Infant mortality (up to 12 months of age) 2. Seizures (definition unreported) (neonatal) 3. Intracranial haemorrhage (definition unreported) (neonatal)
Topp, Langhoff-Roos and Uldall 1997 Denmark	Case control study	(862) Data from Danish cerebral palsy register, 175 singleton CP infants evaluated, controls selected from all preterm singleton live born infants in Eastern Denmark during the same period, matched by gestational age and year of birth Year of birth: 1982-1986 Congenital anomalies NOT excluded	< 37 weeks gestation LMP/ USS in some cases	1 min <7 5 min <7	Cerebral palsy (definition unreported) (age 4-6 years)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Tran, Gray and O'Callaghan 2005 Australia	Case control study	(149) Cohort of infants born at <28 weeks gestation, at single tertiary centre, enrolled in neonatal follow up program. Year of birth: 1989-1996 Exclusions: congenital/ chromosomal anomalies excluded. Multiple pregnancy of triplets or higher.	≥ 24 weeks, <28 weeks gestation LMP/ USS < 20 weeks	1 min <6 5 min <6	Cerebral palsy (persistent abnormality of movement and posture resulting in impairment of function due to non-progressive lesion in immature brain) (age 2 years)
Valentin et al 1993 Sweden	Retrospective cohort	(200) All babies with 1 min Apgar ≤ 8 born at a single centre and 52 randomly selected babies with 1 min Apgar score ≥9. Year of birth 1982 No exclusions reported	Unreported	1 min <7 5 min <7	Neonatal morbidity (neonatal unit admission requiring specific treatment, e.g. Assisted ventilation, IV fluids, death or survival with sequelae)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Walstab et al 2004 Australia	Case control study	(439) Cases included 148 children with moderate or severe CP from Victorian cerebral palsy register born at of 10 hospitals.. Cases matched with 1-2 controls for year of birth, plurality, sex, birth weight, gestation and hospital of birth Year of birth: 1983-1992 Post neonatal cases excluded Congenital anomalies NOT excluded	Unreported	1 min ≤ 3 1 min ≤ 6 5 min ≤ 6	Cerebral palsy (definition and age unreported)
Weinberger et al 2000 USA	Prospective cohort	(852)Neonates with birth weight 500-2000g at 3 hospitals enrolled for follow up Year of birth: 1984-1987 No exclusions reported	23-34 weeks USS <20 weeks	5 min <6	1.Neonatal death (exact age unreported) 2. Respiratory distress syndrome (definition unreported) (neonatal)

Appendix 14. Characteristics of included studies in systematic review of Apgar score and adverse outcomes

Zanini et al 2011 Brazil	Retrospective cohort	(134933) All live births to women living in the State of Rio Grande do Sul, with birth certificates. Data from live birth and death database Year of birth: 2003 Exclusions: missing data, presumed errors of incompatible gestation and birth weight. Congenital anomalies not excluded	>22 weeks gestation Method unreported	1 min <7 5 min <7	Neonatal death (<28 days)

Min= Minutes, mg/dl= miligrams per decilitre, g=grams, USS= ultrasound scan, PVL= periventricular leukomalacia, LMP= last menstrual period, ICD= International Center for disease classification, CSF= cerebrospinal fluid, NICU= neonatal intensive care unit, IQ= intelligence quotient, DQ= Development quotient, CNS= central nervous system, CP= cerebral palsy, SGA= small for gestational age, DSM= diagnostic and statistical manual of mental disorders, SD= standard deviation

Appendix 15. Reference list of included studies in systematic review of Apgar score

Abrams ME, Meredith KS, Kinnard P, Clark RH. Hydrops fetalis: a retrospective review of cases reported to a large national database and identification of risk factors associated with death. *Pediatrics* 2007; 120(1):84-89.

Adamson S, Alessandri L, Badawi N, Burton P, Pemberton P, Stanley F. Predictors of neonatal encephalopathy in full term infants. *BMJ* 1995; 311:598-602.

Aggarwal AK, Kumar P, Chowdary G, Majumdar S, Narang A. Evaluation of renal functions in asphyxiated newborns. *Journal of Tropical Pediatrics* 2005; 51(5): 295-299.

Ajayi O, Nzeh DA. Intraventricular haemorrhage and periventricular leukomalacia in Nigerian infants of very low birth weight. *West African Journal of Medicine* 2003; 22(2):164-166.

Ambalavanan N, Carlo WA. Comparison of the prediction of extremely low birth weight neonatal mortality by regression analysis and by neural networks. *Early Human Development* 2001; 65(2):123-137.

Anyaegbunam A, Fleischer A, Whitty J, Brustman L, Randolph G, Langer O et al. Association between umbilical artery cord pH, five-minute Apgar scores and neonatal outcome. *Gynecologic & Obstetric Investigation* 1991; 32(4):220-223.

Apgar V, James LS. Further observations on the newborn scoring system. *American Journal of Disease in Childhood* 1962 104(4):419-428.

Atkinson D. An evaluation of Apgar scores as predictors of infant mortality. *North Carolina Medical Journal* 1983; 44(1):45-54.

Baenziger O, Moenkhoff M, Morales CG, Waldvogel K, Wolf M, Bucher H et al. Impaired chemical coupling of cerebral blood flow is compatible with intact neurological outcome in neonates with perinatal risk factors. *Biology of the Neonate* 1999; 75(1):9-17.

Bauder FH, von Siebenthal K, Bucher HU. Ultrasonically established cystic periventricular leukomalacia (PVL): incidence and associated factors in Switzerland 1995-1997. *Zeitschrift fur Geburtshilfe und Neonatologie* 2000; 204(2):68-73.

Beeby PJ, Elliott EJ, Henderson-Smart DJ, Rieger ID. Predictive value of umbilical artery pH in preterm infants. *Archives of Disease in Childhood* 1994; 71(2):F93-F96.

Behnke M, Carter R, Hardt N, Eyler F, Cruz A, Resnick M. The relationship of Apgar scores, gestational age, and birthweight to survival

of low birth weight infants. *Am erican Journal of Perinatology* 1987; 4:121-124.

Behnke M, Eyler FD, Carter RL, Hardt NS, Cruz AC, Resnick MB. Predictive Value of Apgar Scores for Developmental Outcome in Premature-Infants. *American Journal of Perinatology* 1989; 6(1):18-21.

Bennett FC, Robinson NM, Sells CJ. Growth and development of infants weighing less than 800 grams at birth. *Pediatrics* 1983; 71(3):319-323.

Berger R, Bender S, Seflow S, Klingmuller V, Kunzel W, Jensen A. Peri/intraventricular haemorrhage: A cranial ultrasound study on 5286 neonates. *European Journal of Obstetrics Gynecology and Reproductive Biology* 1997; 75(2): 191-203.

Brandalise SR, Suguihara CY, Succi RM, Muller RL, Toledo RJ. Apgar score and blood coagulation factors. *Revista Brasileira de Pesquisas Medicas e Biologicas* 1976; 9(1):37-44.

Buchmayer S, Johansson S, Johansson A, Hultman CM, Sparen P, Cnattingius S. Can association between preterm birth and autism be explained by maternal or neonatal morbidity? *Pediatrics* 2009; 124(5):e817-e825.

Camp BW, Broman SH, Nichols PL, Leff M. Maternal and neonatal risk factors for mental retardation: defining the 'at-risk' child. *Early Human Development* 1998; 50(2):159-173.

Casey BM, McIntire DD, Leveno KJ. The continuing value of the Apgar score for the assessment of newborn infants. *New England Journal of Medicine* 2001; 344(7):467-471.

Chandola CA, Robling MR, Peters TJ, Melville-Thomas G, McGuffin P. Pre- and perinatal factors and the risk of subsequent referral for hyperactivity. *Journal of Child Psychology & Psychiatry & Allied Disciplines* 1992; 33(6):1077-1090.

Colburn D, Salzman M. The Apgar scoring system in evaluation of the newborn infant. *New York State J Med* 1960; 60:246-249.

Dalens B, Viallard JL, Raynaud EJ, Dastuge B. CSF levels of lactate and hydroxybutyrate dehydrogenase as indicators of neurological sequelae after neonatal brain damage. *Developmental Medicine & Child Neurology* 1981; 23(2):228-233.

De Almeida MFB, Guinsburg R, Martinez FE, Procianoy RS, Leone CR, Marba STM et al. Perinatal factors associated with early deaths of

preterm infants born in Brazilian Network on Neonatal research centers. *Jornal de Pediatria* 2008; 84(4): 300-307.

den Ouden L, Verloove-Vanhorick SP, van Zeben-van der Aa DM, Brand R, Ruys JH. Neonatal neurological dysfunction in a cohort of very preterm and/or very low birthweight infants--relation to other perinatal factors and outcome at 2 years. *Neuropediatrics* 1990; 21(2):66-71.

Deorari AK, Paul VK, Singh M. Birth asphyxia and neurodevelopmental outcome. *Indian Pediatrics* 1989; 26(8):793-799.

Dijxhoorn MJ, Visser GH, Fidler V, Touwen BC, Huisjes HJ. Apgar score, meconium and acidaemia at birth in relation to neonatal neurological morbidity in term infants. *BJOG* 1986; 93: 217-22.

Drage JS, Berendes H, Fisher PD. The Apgar score and four-year psychological examination performance. In: Perinatal Factors affecting Human Development (Ed Drage JS, Berendes H, Fisher PD) 1969; 85: 222-227.

Drage JS, Kennedy C, Berendes H, Schwarz BK, Weiss W. The Apgar score as an index of infant morbidity. A report from the collaborative study of cerebral palsy. *Developmental Medicine & Child Neurology* 1966; 8(2):141-148.

Drage J, Kennedy C, Schwartz B. The Apgar score as an index of neonatal mortality. *Obstetrics & Gynecology* 1964; 24:222-230.

Echandia CA, Ruiz JG. Low Apgar score and neonatal seizures: neuromotor development at 1 year age. *Colombia Medica* 2006; 37(1):21-30.

Ehrenstein V, Sorensen HT, Pedersen L, Larsen H, Holsteen V, Rothman KJ. Apgar score and hospitalization for epilepsy in childhood: a registry-based cohort study. *Bmc Public Health* 2006; 6: 23

Ekure EN, Ezeaka VC, Iroha E, Egri-Okwaji MTC. Prospective audit of perinatal mortality among inborn babies in a tertiary health center in Lagos, Nigeria. *Nigerian Journal of Clinical Practice* 2011; 14(1):88-94.

Evans N, Hutchinson J, Simpson JM, Donoghue D, Darlow B, Henderson-Smart D et al. Prenatal predictors of mortality in very preterm infants cared for in the Australian and New Zealand Neonatal Network. *Archives of Disease in Childhood Fetal & Neonatal Edition* 2007; 92(1):F34-F40.

Graham EMH. Holcroft CJ, Blakemore KJ. Evidence of intrapartum hypoxia ischemia is not present in the majority of cases of neonatal

seizures. *Journal of Maternal-Fetal and Neonatal Medicine* 2002; 12(2):123-6.

Grether JK, Nelson KB, Stanley-Emery E3, Cummins SK. Prenatal and perinatal factors and cerebral palsy in very low birthweight infants. *Journal of Pediatrics* 1996; 128:407-414.

Heller G, Schnell RR, Misselwitz B, Schmidt S, . Umbilical blood pH, Apgar scores, and early neonatal mortality. *Zeitschrift fur Geburtshilfe und Neonatologie* 2003; 207(3):84-89.

Heuchan A, Evans N, Henderson Smart D. Perinatal risk factors for major intraventricular haemorrhage in the Australian and New Zealand Neonatal Network. *Arch Dis Childhood* 2002; 86:F86-F90.

Holst K, Andersen E, Philip J, Henningsen I. Antenatal and perinatal conditions correlated to handicap among 4-year-old children. *American Journal of Perinatol*ogy 1989; 6(2):258-267.

Iijima S, Arai H, Ozawa Y, Kawase Y, Uga N. Clinical patterns in extremely preterm (22 to 24 weeks of gestation) infants in relation to survival time and prognosis. *American Journal of Perinatology* 2009; 26(6):399-406.

Ikonen RS. The Apgar scoring of newborn infants and its relation to neonatal mortality. *Annales Paediatriae Fenniae* 1967; 13(4):111-114.

Ikonen RS. Apgar scoring and neonatal morbidity in full-sized newborn infants. *Annals of Clinical Research* 1973; 5(6):380-384.

Ikonen RS, Lauslahti K. Apgar scoring and neonatal morbidity in infants weighing 1000-1990 g. *Annals of Clinical Research* 1974; 6(3):161-164.

Indredavik MS, Vik T, Evensen KAI, Skranes J, Taraldsen G, Brubakk AM. Perinatal Risk and Psychiatric Outcome in Adolescents Born Preterm With Very Low Birth Weight or Term Small for Gestational Age. *Journal of Developmental and Behavioral Pediatrics* 2010; 31(4):286-294.

Ishikawa TK, Kishi S, Inukai K, Kono C, Kitoh H, Awaya A et al. Subsequent epilepsy in very-low-birthweight infants: A long-term follow-up study from birth. *Epilepsia* 1995; 36(5):435-439.

Issel EP, Eggers H, Plath C, Towe J, Voigt M. The apgar value of the newborn and its prognostic value for the course of the neonatal period. *Zentralblatt fur Gynakologie* 1976; 98(26):1618-1625.

Jacobsson BH, Hagberg G, Hagberg B, Ladfors L, Niklasson A, Hagberg H. Cerebral palsy in preterm infants: A population-based case-control

study of antenatal and intrapartal risk factors. *Acta Paediatrica, International Journal of Paediatrics* 2002; 91(8):946-951.

Jennett RJ, Warford HS, Kreinick C, Waterkotte GW. Apgar index: a statistical tool. *American Journal of Obstetrics and Gynecology* 1981; 140(2):206-212.

Johnson LH, Sivieri E, Bhutani V. Apgar score influences neurologic outcome in infants with hyperbilirubinemia. *Pediatric Research* 1999; 45(4):1189.

Kato EHY, Yamada H, Matsumoto Y, Hattori S, Makinoda S, Fujimoto S. Relation between perinatal factors and outcome of very low birth weight infants. *Journal of Perinatal Medicine* 1996; 24(6):677-686.

Krebs L, Langhoff-Roos J, Thorngren-Jerneck K. Long-term outcome in term breech infants with low Apgar score--a population-based follow-up. *European Journal of Obstetrics, Gynecology, & Reproductive Biology* 2001; 100(1):5-8.

Lee HC, Subeh M, Gould JB. Low Apgar score and mortality in extremely preterm neonates born in the United States. *Acta Paediatrica* 2010; 99(12):1785-1789.

Lie KKG, Groholt E-K, Eskild A. Association of cerebral palsy with Apgar score in low and normal birthweight infants: population based cohort study. *BMJ* 2010: 341:c4990

Luthy DA, Shy KK, Strickland D, Wilson J, Bennett FC, Brown ZA et al. Status of infants at birth and risk for adverse neonatal events and long-term sequelae: a study in low birth weight infants. *American Journal of Obstetrics & Gynecology* 1987; 157(3):676-679.

Mahon BE, Ehrenstein V, Norgaard M, Pedersen L, Rothman KJ, Sorensen HT et al. Perinatal risk factors for hospitalization for pneumococcal disease in childhood: a population-based cohort study. *Pediatrics* 2007; 119(4):e804-e812.

Minchom PN, Niswander K, Chalmers I, Dauncey M, Newcombe R, Elbourne D et al. Antecedents and outcome of very early neonatal seizures in infants born at or after term. *British Journal of Obstetrics and Gynaecology* 1987; 94(5): 431-439.

Misra PK, Srivastava N, Malik GK, Kapoor RK, Srivastava KL, Rastogi S et al. Outcome in relation to Apgar score in term neonates. *Indian Pediatrics* 1994; 31(10):1215-1218.

Appendix 15. Reference list of included studies in systematic review of Apgar score

Moro M, Figueras-Aloy J, Fernandez C, Domenech E, Jimenez R, Perez-Rodriguez J et al. Mortality for Newborns of birthweight less than 1500 g in Spanish neonatal units (2002-2005). *American Journal of Perinatology* 2007; 24(10):593-601.

Moster D, Lie RT, Irgens LM, Bjerkedal T, Markestad T. The association of Apgar score with subsequent death and cerebral palsy: A population-based study in term infants. *Journal of Pediatrics* 2001; 138(6):798-803.

Moster D, Lie RT, Markestad T. Joint association of Apgar scores and early neonatal symptoms with minor disabilities at school age. *Archives of Disease in Childhood Fetal & Neonatal Edition* 2002; 86(1):F16-F21.

Moura DR, Costa JC, Santos IS, Barros AJ, Matijasevich A, Halpern R et al. Natural history of suspected developmental delay between 12 and 24 months of age in the 2004 Pelotas birth cohort. *Journal of Paediatrics & Child Health* 2010; 46(6):329-336.

Murphy DJ, Sellers S, MacKenzie IZ, Yudkin PL, Johnson AM. Case-control study of antenatal and intrapartum risk factors for cerebral palsy in very preterm singleton babies. *Lancet* 1995; 346(8988):1449-1454.

Myers SA, Paton JB, Fisher DE. The effect of initial Apgar score on the birthweight-specific survival of the very low-birthweight infant. *American Journal of Perinatology* 1987; 4(4):288-292.

Nelson KB, Ellenberg JH. Apgar scores as predictors of chronic neurologic disability. *Pediatrics* 1981; 68:36-44.

Oain P, Skramm I, Hannisdalm E, Bjoro K. Breech delivery: an obstetrical analysis. *Acta Obstetrica Gynecologica Scandinavia* 1988; 67:75-79.

Obwegeser R, Bohm R, Gruber W. The Significance of Divergent Umbilical-Cord Ph Levels and Apgar-Scores of Newborns. *Zeitschrift fur Geburtshilfe und Perinatologie* 1993; 197(2):59-64.

Paludetto RC, Corchia C, Maltese M, De Curtis M, Taurisano I, Cipollone A et al. Apgar score, neonatal neurlogical signs and late cerebral damage in full term infants. *Rivista Italiana di Pediatria* 1981; 7(1):29-36.

Patterson CC, Carson DJ, Hadden DR, Waugh NR, Cole SK. A case control investigation of perinatal risk factors for childhood IDDM in Northern Ireland and Scotland. *Diabetes Care* 1994; 17:376-381.

Paul RH, Koh KS, Monfared AH. Obstetric factors influencing outcome in infants weighing from 1,001 to 1,500 grams. *American Journal of Obstetrics and Gynecology* 1979; 133(5):503-508.

Appendix 15. Reference list of included studies in systematic review of Apgar score

Perlman JM, Risser R. Can asphyxiated infants at risk for neonatal seizures be rapidly identified by current high-risk markers? *Pediatrics* 1996; 97(4):456-462.

Robertson L, Harrild K. Maternal and neonatal risk factors for childhood type 1 diabetes: a matched case-control study. *Bmc Public Health* 2010; 10.

Ruth VJ, Raivio KO. Perinatal brain damage: predictive value of metabolic acidosis and the APGAR score. *British Medical Journal* 1988; 297: 24-27.

Sanders AE, Slade GD. Apgar score and dental caries risk in the primary dentition of five year olds. *Australian Dental Journal* 2010; 55(3):260-267.

Schmidt LS, Schuz J, Lahteenmaki P, Trager C, Stokland T, Gustafson G et al. Fetal Growth, Preterm Birth, Neonatal Stress and Risk for CNS Tumors in Children: A Nordic Population- and Register-Based Case-Control Study. *Cancer Epidemiology Biomarkers & Prevention* 2010; 19(4):1042-1052.

Seidman DS, Paz I, Laor A, Gale R, Stevenson DK, Danon YL. Apgar Scores and Cognitive Performance at 17 Years of Age. *Obstetrics and Gynecology* 1991; 77(6):875-878.

Serenius F, Ewald U, Farooqi A, Holmgren PA, Hakansson S, Sedin G et al. Short-term outcome after active perinatal management at 23-25 weeks of gestation. A study from two Swedish tertiary care centres. Part 2: infant survival. *Acta Paediatrica* 2004; 93(8):1081-1089.

Shankaran S, Fanaroff AA, Wright LL, Stevenson DK, Donovan EF, Ehrenkranz RA et al. Risk factors for early death among extremely low-birth-weight infants. *American Journal of Obstetrics and Gynecology* 2002; 186(4):796-802.

Soleimani F, Vameghi R, Biglarian A, Daneshmandan N. Risk Factors Associated with Cerebral Palsy in Children Born in Eastern and Northern Districts of Tehran. *Iranian Red Crescent Medical Journal* 2010; 12(4):428-433.

Stelmach TP, Pisarev H, Talvik T. Ante- and perinatal factors for cerebral palsy: Case-control study in Estonia. *Journal of Child Neurology* 2005; 20(8): 654-661.

Tejani N, Verma U. Correlation of Apgar scores and umbilical artery acid-base status to mortality and morbidity in the low birth weight neonate. *Obstetrics & Gynecology* 1989; 73:597-600.

Appendix 15. Reference list of included studies in systematic review of Apgar score

Thorngren-Jerneck K, Herbst A. Low 5-minute Apgar score: a population-based register study of 1 million term births. *Obstetrics & Gynecology* 2001; 98(1):65-70.

Topp M, Langhoff-Roos J, Uldall P. Preterm birth and cerebral palsy. Predictive value of pregnancy complications, mode of delivery, and Apgar scores. *Acta Obstetricia et Gynecologica Scandinavica* 1997; 76(9):843-848.

Tran U, Gray P, O'Callaghan M. Neonatal antecedents for cerebral palsy in extremely preterm babies and interaction with maternal factors. *Early Human Development* 2005; 81:555-561.

Valentin L, Ekman G, Isberg PE, Polberger S, Marsal K. Clinical evaluation of the fetus and neonate. Relation between intra-partum cardiotocography, Apgar score, cord blood acid-base status and neonatal morbidity. *Archives of Gynecology & Obstetrics* 1993; 253(2):103-115.

Walstab JE, Bell RJ, Reddihough DS, Brennecke SP, Bessell CK, Beischer NA et al. Factors identified during the neonatal period associated with risk of cerebral palsy. *Australian & New Zealand Journal of Obstetrics & Gynaecology* 2004; 44(4):342-346.

Weinberger B, Anwar M, Hegyi T, Hiatt M, Koons A, Paneth N et al. Antecedents and neonatal consequences of low Apgar scores in preterm newborns: a population study. *Archives of Pediatrics & Adolescent Medicine* 2000; 154(3):294-300.

Zammit S, Odd D, Horwood J, Thompson A, Thomas K, Menezes P et al. Investigating whether adverse prenatal and perinatal events are associated with non-clinical psychotic symptoms at age 12 years in the ALSPAC birth cohort. *Psychological Medicine* 2009; 39(9):1457-1467.

Zanini RR, de Moraes AB, Giugliani ERJ, Riboldi J. Contextual determinants of neonatal mortality using two analysis methods, Rio Grande do Sul, Brazil. *Revista de Saude Publica* 2011; 45(1):79-89.

Appendix 16. Formulae for calculating post-test probability from likelihood ratios

The formula is as follows:[178]

- Pretest odds = (Pretest probability / (1 - Pretest probability)
- Posttest odds = Pretest odds * Likelihood ratio

In equation above, *positive post-test probability* is calculated using the *likelihood ratio positive*, and the *negative post-test probability* is calculated using the *likelihood ratio negative*.

- Posttest probability = Posttest odds / (Posttest odds + 1)

Appendix 17. Medline search strategy for cost-effectiveness analyses of neonatal hypothermia to prevent cerebral palsy

1. cost effectiveness.mp. or Cost-Benefit Analysis/
2. Hypothermia, Induced/ or neonatal cooling.mp.
3. cerebral palsy.mp. or Cerebral Palsy/
4. neurodevelopmental delay.mp.
5. 1 and 2 and 3 and 4
6. 1 and 2 and 3
7. 1 and 2

Appendix 18. Calculations of costs for decision-analytic model not obtained directly from the literature

Cost of test

Cost of umbilical cord pH test on point of care analyser= £2.42 (University of Nottingham)

Cost of midwife x 1hr patient contact (including qualifications)= £97[174]

Assume it takes 10 minutes for midwife to perform test

Cost of performing umbilical cord pH test= (97/6)+ 2.42= £18.59

Community costs to age 18 months of a non-encephalopathic baby

Average number of GP consultations per person per year= 6.5 in 0-4 year age group (2008/2009)[173]

6.5 x1.5= 9.75 consultations in 18 month period

Cost of 10 minute GP consultation= £36[174]

£36 x 9.75 = £351

Four visits to practice nurse for routine immunisations (15 mins per visit, 1 hour total)

Cost of 1 hour practice nurse time= £39[174]

Child health surveillance, cost for first three visits (to 18 months)= £10.74 + £6.82 + £9.44 = £27.00 (1997 costs) Inflated to 2011 costs= £50

Appendix 18. Calculations of costs for decision-analytic model not obtained directly from the literature

£351 +£39 +£50 = **£440 total community costs of a non-encephalopathic baby to 18 months of age**

Inpatient costs at time of delivery for a non-encephalopathic baby

0.1% of obstetric patients would have a neonate with multiple minor diagnoses, and 0.7% would have a neonate with one minor diagnosis of all obstetric unit admissions.[171] When restricted to delivery events only (51.2%), this translated to 0.2% and 1.4% respectively.

Mean cost of neonate with multiple minor diagnoses= £716 (2005/2006)[171]

Mean cost of neonate with one minor diagnosis= £1,091(2005/2006)[171]

0.002 x 1,091= £2.18

0.014 x 763= £ 10.68

Inpatient cost at time of delivery= £2.18 +£10.68= £12.86, inflated to 2010/11 costs = £14.74

Hospital costs of a non-encephalopathic baby to age 18 months

Hospital admission rate (2007/2008) 0-5 year age group =128.9 per 1000 population[172]

=0.13% x1.5 (18 months) = 0.20

Cost of hospital readmission (2006/2007)= £480[163]

Appendix 18. Calculations of costs for decision-analytic model not obtained directly from the literature

Assume stay one day

0.20 x £480= £96, inflated to 2010/2011 = **£106.08 hospital costs to age 18 months**

Total hospital costs: £14.74 + £106.08 = £120.82

In all cases, cost inflations were performed according to the following formula:

((Current price index/ original price index) – 1) x original cost = added cost in 2010/2011

E.g. 2006/2007 index = 249.8[174]

2010/2011= 276[174]

276/249.8-1= 0.105

0.105 x£96= £10.08

Therefore 2010/2011 cost =£96 +£10.08=£106.08

Appendix 19. List of variable definitions for decision-analytic model

Table 1. Distributions

NAME	DESCRIPTION	TYPE	PARAMETERS	Estimated Value
speccordpH_7	Specificity cord pH threshold <7.00	Beta	subtype: 2, alpha: 137.8, beta: 62.2	0.689
senscordpH_7	Sensitivity cord pH threshold <7.00	Beta	subtype: 2, alpha: 3.75, beta: 2.25	0.625
speccordpH_710	Specificity cord pH threshold <7.10	Beta	subtype: 2, alpha: 310.08, beta: 29.92	0.912
senscordpH_710	Sensitivity cord pH threshold <7.10	Beta	subtype: 2, alpha: 6.042, beta: 31.958	0.159
RRCPcooling_7	Relative risk of neonatal cooling for cerebral palsy in a population with pH <7.00	LogNormal	umeanoflogs: -0.371, sigmastddevoflogs: 0.127	0.69563

Appendix 19. List of variable definitions for decision-analytic model

RRcoolingthresh_710	Relative risk of neonatal cooling for cerebral palsy in a population with pH <7.10	LogNormal	umeanoflogs: 0.981, sigmastddevoflogs: 1.041	4.58523
CP_prevalence	Prevalence of cerebral palsy	Beta	subtype: 2, alpha: 2, beta: 998	0.002
cost_cooling	Cost of neonatal cooling including equipment	Gamma	alpha: 3.2, lambda: 0.000489	6543.967
cost_hosp_cool	Hospital costs for initial inpatient stay per cooled infant	Gamma	alpha: 122, lambda: 0.00875	13942.86
cost_hospread_cooled	Hospital readmission costs in the first year of life per cooled infant	Gamma	alpha: 16.9, lambda: 0.0141	1198.582
cost_transfer_cool	Neonatal transfer costs per cooled infant	Gamma	alpha: 127, lambda: 0.7840	161.9898
cost_comm6_cool	Community care costs per cooled infant until 6 months of age	Gamma	alpha: 106, lambda: 0.14077	753.0014

Appendix 19. List of variable definitions for decision-analytic model

cost_comm612_cool	Community care costs per cooled infant from 6-12 months of age	Gamma	alpha: 33.5, lambda: 0.06893	486.0003
cost_inpatient_noncool	Hospital costs for initial inpatient stay per non-cooled infant (with encephalopathy)	Gamma	alpha: 118, lambda: 0.00839	14064.36
cost_hospread_noncool	Hospital readmission costs in the first year of life per non-cooled infant (with encephalopathy)	Gamma	alpha: 10, lambda: .003775	2649.007
cost_trans_noncool	Neonatal transfer costs per non-cooled infant (with encephalopathy)	Gamma	alpha: 110, lambda: .6918	159.0055
cost_comm6_noncool	Community care costs per non-cooled infant until 6 months of age (with encephalopathy)	Gamma	alpha: 66, lambda: .08365	789.0018

Appendix 19. List of variable definitions for decision-analytic model

cost_comm612_noncool	Community care costs per non-cooled infant from 6-12 months of age (with encephalopathy)	Gamma	alpha: 23.5, lambda: .031085	755.9916
Cost_CP_comm_cool	Cost of community care per cooled infant, who is diagnosed with cerebral palsy, from 12-18 months	Gamma	alpha: 12.6, lambda: .03378	373.0018
cost_CP_cool_hosp	Inpatient costs care per cooled infant, who is diagnosed with cerebral palsy, from 12-18 months	Gamma	alpha: 6.54, lambda: .00375	1744

Appendix 19. List of variable definitions for decision-analytic model

cost_CPcomm_noncool	Cost of community care per non-cooled infant, who is diagnosed with cerebral palsy, from 12-18 months	Gamma	alpha: 16, lambda: .031558	507.003
cost_CPhosp_noncool	Inpatient costs of care per non-cooled infant, who is diagnosed with cerebral palsy, from 12-18 months	Gamma	alpha: 5.24, lambda: .004292	1220.876
cost_noCPinpt_cool	Inpatient costs of care per cooled infant, who does not develop cerebral palsy, from 12-18 months	Gamma	alpha: 3.2, lambda: .022378	142.9976

Appendix 19. List of variable definitions for decision-analytic model

cost_comm_noCPcool	Costs of community care per cooled infant, who does not develop cerebral palsy, from 12-18 months	Gamma	alpha: 12.5, lambda: .052966	236.0005
cost_hospnoCP_noncool	Inpatient costs of care per cooled infant, who does not develop cerebral palsy, from 12-18 months	Gamma	alpha: 3.2, lambda: .022378	142.9976
Cost_commnoCP_noncool	Costs of community care per cooled infant, who does not develop cerebral palsy, from 12-18 months	Gamma	alpha: 12.7, lambda: .053814	235.9981

Appendix 19. List of variable definitions for decision-analytic model

Table 2. Variables

NAME	DESCRIPTION	DEFINED AT	FORMULA	VALUE
cordph7	Probability of a positive cord pH (<7.00) test	Node1: Neonates	(senscordpH_7*CP_prevalence)+((1-speccordpH_7)*(1-CP_prevalence))	0.311627999999999
Pcordph71	Probability of a positive cord pH (<7.10) test	Node1: Neonates	(senscordpH_710*CP_prevalence)+((1-speccordpH_710)*(1-CP_prevalence))	0.0881420000000008
PPVCordPh7	Positive predictive value of cord pH <7.00	Node1: Neonates	(senscordpH_7*CP_prevalence)/((senscordpH_7*CP_prevalence)+((1-speccordpH_7)*(1-CP_prevalence)))	0.004011192832479753
npvcordph7	Negative predictive value of cord pH <7.00	Node1: Neonates	(speccordpH_7*(1-CP_prevalence))/(((1-senscordpH_7)*CP_prevalence)+((speccordpH_7)*(1-CP_prevalence)))	0.9989104728257395
ppVcordpH71	Positive predictive value of cord pH <7.10	Node1: Neonates	(senscordpH_710*CP_prevalence)/((senscordpH_710*CP_prevalence)+((1-speccordpH_710)*(1-CP_prevalence)))	0.0036078146627033622

Appendix 19. List of variable definitions for decision-analytic model

NPVcordpH71	Negative predictive value of cord pH <7.10	Node1: Neonates	(speccordpH_710*(1-CP_prevalence))/(((1-senscordpH_710)*CP_prevalence)+((speccordpH_710)*(1-CP_prevalence)))	0.9981554145491952
cost_comm_well	Community costs of a non-encephalopathic infant, who does not develop cerebral palsy, until 18 months of age	Node1: Neonates	440.00	440.0
cost_hosp_well	Hospital costs of a non-encephalopathic infant, who does not develop cerebral palsy, until 18 months of age, including neonatal period	Node1: Neonates	120.82	120.82
cost_cordpH	Cost of performing cord pH test	Node1: Neonates	18.59	18.59

(1) Medical research council. MRC Research priorities. 2012. http://www.mrc.ac.uk/About/Strategy/StrategicPlan2009-2014/index.htm

(2) Parkes J, Hill N. The needs of children and young people with cerebral palsy. *Paediatric Nursing* 2010; 22(4):14-18.

(3) Office of National Statistics. Births and deaths in England and Wales, 2010. 2011. http://www.ons.gov.uk/ons/rel/vsob1/birth-summary-tables--england-and-wales/2010/index.html

(4) Delivoria-Papadopoulos M, McGowan JE. Oxygen transport and delivery. In: Polin RA, Fox WW, Abman SH, editors. Fetal and neonatal physiology. fourth ed. Philadelphia: Saunders, Elesvier; 2011. 970-979.

(5) Hall JE. Acid base regulation. In: Hall J, editor. Textbook of medical physiology. 12th ed. Philadelphia: Saunders; 2011. 379-412.

(6) Nageotte MP, Gilstrap LC. Intrapartum fetal surveillance. In: Creasy RK, Resnik R, Iams JD, Lockwood CJ, Moore TR, editors. Creasy and Resnik's Maternal- fetal medicine. 6th ed. Philadelphia: Saunders; 2009. 397-418.

(7) Fahey J, King TL, Fahey J, King TL. Intrauterine asphyxia: clinical implications for providers of intrapartum care. *Journal of Midwifery & Women's Health* 2005; 50(6):498-506.

(8) Goldaber KG. Pathological fetal acidemia. *Obstetrics & Gynecology* 1991; 78: 1103-1107.

(9) Ross MG. Gala, R. Use of umbilical artery base excess: Algorithm for the timing of hypoxic injury. *American Journal of Obstetrics and Gynecology* 2002; 187(1): 1-9.

(10) Giussani DA, Gardner DS. Intrauterine hypoxaemia and cardiovascular development. In: Langley-Evan SC, editor. Fetal nutrition and adult disease. first ed. Wallingford, Oxfordshire: CABI Publishing; 2004. 55-85.

(11) Volpe JJ. Perinatal brain injury: from pathogenesis to neuroprotection. *Mental Retardation and Developmental Disabilities Reseach Reviews* 2001; 7(1):56-64.

(12) Levene ML, Kornberg J, Williams TH. The incidence and severity of post-asphyxial encephalopathy in full-term infants. *Early Human Development* 1985; 11: 21-26.

(13) Vannucci RC. Current and potentially new management strategies for perinatal hypoxic ischaemic encephalopathy. *Pediatrics* 1990; 85:961-968.

(14) Herbst A, Thorngren-Jerneck K, Wu L, Ingemarsson I. Different types of acid-base changes at birth, fetal heart rate patterns, and infant outcome at 4 years of age. *Acta Obstetricia et Gynecologica Scandinavica* 1997; 76(10):953-958.

(15) Kato EHY, Yamada H, Matsumoto Y, Hattori S, Makinoda S, Fujimoto S. Relation between perinatal factors and outcome of very low birth weight infants. *Journal of Perinatal Medicine* 1996; 24(6):677-686.

(16) Hossain MA. Hypoxic ischemic injury in neonatal brain: involvement of a novel neuronal molecule in neuronal cell death and potential target for neuroprotection. *International Journal of Developmental Neuroscience* 2008; 26:93-101.

(17) Rees S, Inder T. Fetal and neonatal origins of altered brain development. *Early Human Development* 2005; 81:753-761.

(18) Luciana M. Cognitive development in children born preterm: Implications for theories of brain plasticity following early injury. *Development in psychopathology* 2003; 13:1017-1047.

(19) Huch R, Huch A. Maternal-fetal acid base balance and blood gas measurement. In: Beard RW, Nathanielsz PW, editors. Fetal physiology and medicine. New York: Marcel Dekker; 1984. 713.

(20) Riley RJ, Johnson JWC. Collecting and analyzing cord blood gases. *Clinical Obstetrics & Gynecology* 1993; 36:13-23.

(21) Hall JE. Fetal and neonatal physiology. In: Hall JE, Guyton AC, editors. Textbook of Medical Physiology. 12th ed. Philadelphia: Saunders; 2011. 1019-1028.

(22) Saugstad OD. Physiology of resuscitation. In: Polin RA, Fox WW, Abman SH, editors. Fetal and neonatal physiology. 4th ed. Philadelphia: Saunders; 2011. 846-863.

(23) Noori S, Friedlich PS, Seri I. Pathophysiology of shock in the fetus and neonate. In: Polin RA, Fox WW, Abman SH, editors. Fetal and neonatal physiology. 4th ed. Philadelphia: Saunders; 2011. 853-863.

(24) Rennie JM. Neurological problems in the newborn. In: Rennie JM, editor. Robertson's textbook of neonatology. 4th ed. Elsevier; 2005. 1093.

(25) Laugel V, Cossee M, Matis J, de Saint-Martin A, Echaniz-Laguna A, Mandel J-L et al. Diagnostic approach to neonatal hypotonia: retrospective study on 144 neonates. *European Journal of Pediatrics* 2008; 167:517-523.

(26) Oberlander TF, Misri S, Fitzgerald CE, Kostaras X, Rurak D, Riggs W. Pharmacologic factors associated with transient neonatal symptoms following prenatal psychotropic medication exposure. *Journal of Clinical Psychiatry* 2004; 65(2):230-237.

(27) Cunningham FG, Leveno KJ, Bloom SL, et al. Fetal growth disorders. In: Cunningham FG, editor. Williams Obstetrics. 23rd ed. McGraw Hill; 2010. 842.

(28) Gluckman PD, Pinal CS. Regulation of fetal growth by the somatotrophic axis. *The Journal of Nutrition* 2003; 133:1741S-1746S.

(29) Oliver MH, Harding JE, Breier BH, Gluckman PD. Fetal insulin-like growth factor (IGF)-1 and (IGF)-2 are regulated differently by glucose or insulin in the sheep fetus. *Reproduction Fertility and Development* 1994; 8:167-172.

(30) Brodsky D, Christou H. Current concepts in intrauterine growth restriction. *Journal of Intensive Care Medicine* 2004; 19(6):307-319.

(31) Khong T, De Wolf F, Robertson W, Brosens I. Inadequate maternal vascular response to placentation in pregnancies complicated by pre-eclampsia and small for gestational age. *BJOG* 1986; 93:1049-1059.

(32) Alberry M, Soothill P. Management of fetal growth restriction. *Archives of Disease in Childhood: Fetal and Neonatal Edition* 2007; 92:F62-F67.

(33) Bukowski R. Fetal Growth Potential and Pregnancy Outcome. *Seminars in Perinatology* 2004; 28(1): 51-58.

(34) Sheridan C. Intrauterine growth restriction. *Australian Family Physician* 2005; 34(9): 717-23.

(35) Jacobsson BA, Ahlin K, Francis A, Hagberg G, Hagberg H, Gardosi J. Cerebral palsy and restricted growth status at birth: Population-based case-control study. *BJOG: An International Journal of Obstetrics and Gynaecology* 2008; 115(10): 1250-1255.

(36) Jelliffe-Pawlowski LLH, Hansen, R.L. Neurodevelopmental outcome at 8 months and 4 years among infants born full-term

small-for-gestational-age. *Journal of Perinatology* 2004; 24(8): 505-514.

(37) Huxley RR, Shiell AW, Law CM. The role of size at birth and postnatal catch-up growth in determining systolic blood pressure: A systematic review of the literature. *Journal of Hypertension* 2000; 18(7): 815-831.

(38) Newsome CA, Shiell AW, Fall CH, Phillips DI, Shier R, Law CM et al. Is birth weight related to later glucose and insulin metabolism?--A systematic review. *Diabetic Medicine* 2003; 20(5):339-348.

(39) Barker DJ, Osmond C. Infant mortality, childhood nutrition and ischaemic heart disease in England and Wales. *Lancet* 1986; 1:1077-1081.

(40) Joss-Moore LA, Lane RH. The developmental origins of adult disease. *Current Opinion in Pediatrics* 2009; 21:230-234.

(41) Nuyt AM, Szyf M. Developmental programming through epigenetic changes. *Circulation Research* 2007; 100(4): 452-5.

(42) Leon DA, Lithell HO, Vagero D, Koupilova I, Mohsen R, Berglund L et al. Reduced fetal growth rate and increased risk of death from ischaemic heart disease: cohort study of 15 000 Swedish men and women born 1915-29. *BMJ* 1998; 317(7153):241-245.

(43) Martyn CN, Barker DJP, Osmond C. Mothers' pelvic size, fetal growth, and death from stroke and coronary heart disease in men in the UK. *Lancet* 1996; 348(9037):1264-1268.

(44) Rich-Edwards J, Stampfer MJ, Manson J, Rosner B, Hankinson SE, Colditz GA et al. Birth weight and risk of cardiovascular disease in a cohort of women followed up since 1976. *BMJ* 1997; 315:396.

(45) Law CM, Shiell AW. Is blood pressure inversely related to birth weight? The strength of evidence from a systematic review of the literature. *Journal of Hypertension* 1996; 14(8):935-941.

(46) Nobili V, Alisi A, Panera N, Agostoni C. Low birth weight and catch-up-growth associated with metabolic syndrome: A ten year systematic review. *Pediatric Endocrinology Reviews* 2008; 6(2):241-7.

(47) Andersson SW, Lapidus L, Niklasson A, Hallberg L, Bengtsson C, Hulthen L. Blood pressure and hypertension in middle-aged women in relation to weight and length at birth: A follow-up study. *Journal of Hypertension* 2000; 18(12): 1753-61.

(48) Da Silveira VMF, Horta BL. Birth weight and metabolic syndrome in adults: meta-analysis. *Revista de Saude Publica* 2008; 42(1):10-18.

(49) Catalano PM, Farrell K, Thomas A, Huston-Presley L, Mencin P, de Mouzon SH et al. Perinatal risk factors for childhood obesity and metabolic dysregulation. *American Journal of Clinical Nutrition* 2009; 90(5):1303-1313.

(50) Libby G, Mcewan SR, Morris AD, Belch JJF. No difference in the association between birth weight and total cholesterol for males and females. A SHARP (Scottish Heart and Arterial Disease Risk Prevention) study. *Vascular Medicine* 2008; 13(4):271-274.

(51) Libby GM. Birth Weight Does Not Predict Blood Pressure in a Young Working Population: A Sharp (Scottish Heart and Arterial Disease Risk Prevention) Study. *Annals of Epidemiology* 2008; 18(4):Apr.

(52) Wilson J. The Barker hypothesis - An analysis. *Australian & New Zealand Journal of Obstetrics & Gynaecology* 1999; 39(1):1-7.

(53) Davey Smith G, Sterne JAC, Tynelius P, Rasmussen F. Birth characteristics of offspring and parental diabetes: evidence for the fetal insulin hypothesis. *Journal of Epidemiology & Community Health* 2004; 58:126-128.

(54) Eichenwald EC, Stark AR. Management and outcomes of very low birth weight. *New England Journal of Medicine* 2008; 358(16):1700-1711.

(55) Miller MJ, Martin RJ. Pathophysiology of apnea of prematurity. In: Polin RA, Fox WW, Abman SH, editors. Fetal and neonatal physiology. 4th ed. Philadelphia: Saunders; 2011. 998-1011.

(56) National Collaborating centre for womens' and childrens' health (NCC-WCH). Intrapartum care of healthy women and their babies during childbirth. [First]. 2008. National Institute for Clinical Excellence.http://guidance.nice.org.uk/CG55/NICEGuidance/pdf/English

(57) Malin GL, Kigozi P, Budge H. Umbilical cord pH testing and neonatal cooling: Survey of UK neonatal unit practice. 2012. (Unpublished)

(58) Trent Perinatal Network. Neonatal encephalopathy guideline. [2]. 2009. http://www.trentperinatal.nhs.uk/index.asp?pgid=10219

(59) Hall D, Williams J, Elliman D. The child health surveillance handbook. 3rd ed. Radcliffe Publishing; 2009.

(60) Gardosi J, Clausson B. The value of customised centiles in assessing perinatal mortality risk associated with parity and maternal size. *BJOG: An International Journal of Obstetrics & Gynaecology* 2009;1-8.

(61) American Academy of Pediatrics, Committee on Fetus and Newborn, American College of Obstetricians and Gynecologists, Committee on Obstetric Practice. The Apgar score. *Advances in Neonatal Care* 2006; 6(4):220-223.

(62) Apgar V. The newborn (Apgar) scoring system. Reflections and advice. *Pediatric Clinics of North America* 1966; 13(3):645-650.

(63) Potts NC. Assessment at birth and use of the Apgar score: an analysis of the limitations. *Journal of Neonatal Nursing* 2003; 9(3):76-80.

(64) Jepson HA, Talashek ML, Tichy AM. The Apgar score: evolution, limitations, and scoring guidelines. *Birth: Issues in Perinatal Care* 1991; 18(2):83-92.

(65) Richardson DK, Corcoran JD, Escobar GJ, Lee SK. SNAP-II and SNAPPE-II: simplified newborn illness severity and mortality risk scores. *Journal of Pediatrics* 2001; 138: 92-100.

(66) Parry G, Tucker J, Tarnow-Mordi W. CRIB II: an update of the clinical risk index for babies score. *Lancet* 2003; 361:1789-1791.

(67) Heimstad R, Skogvoll E, Mattsson LA, Johansen OJ, Eik-Nes SH, Salvesen KA et al. Induction of labor or serial antenatal fetal monitoring in postterm pregnancy: a randomized controlled trial. *Obstetrics & Gynecology* 2007; 109(3):609-617.

(68) Shennan AH, Smith R, Browne D, Edmonds DK, Morgan B. The elective use of oxytocin infusion during labour in nulliparous women using epidural analgesia: a randomised double-blind placebo-controlled trial. *International Journal of Obstetric Anesthesia* 1995; 4(2):78-81.

(69) Cluett ER, Pickering RM, Getliffe K, St George Saunders NJ. Randomised controlled trial of labouring in water compared with standard of augmentation for management of dystocia in first stage of labour. *BMJ* 2004; 328(7435):314.

(70) Hansen SL, Clark SL, Foster JC. Active pushing versus passive fetal descent in the second stage of labor: a randomized controlled trial. *Obstetrics & Gynecology* 2002; 99(1):29-34.

(71) MacLennan A. A template for defining a causal relation between acute intrapartum events and cerebral palsy: international consensus statement. *BMJ* 1999; 319:1054-1059.

(72) Dear P, Newell S, Rosenbloom L, Rennie JM, MacLennan A. Establishing probable cause in cerebral palsy. *BMJ* 2000; 320:1075.

(73) Dijxhoorn MJ, Visser GH, Huisjes HJ, Fidler V, Touwen BC. The relation between umbilical pH values and neonatal neurological morbidity in full term appropriate-for-dates infants. *Early Human Development* 1985; 11(1):33-42.

(74) Van den Berg PPN. Neonatal complications in newborns with an umbilical artery pH <7.00. *American Journal of Obstetrics and Gynecoogyl* 1996; 175(5): 1152-7.

(75) Wildschut J, Feron FJ, Hendriksen JG, van HM, Gavilanes-Jiminez DW, Hadders-Algra M et al. Acid-base status at birth, spontaneous motor behaviour at term and 3 months and neurodevelopmental outcome at age 4 years in full-term infants. *Early Human Development* 2005; 81(6):535-544.

(76) Gaudier FL, Goldenberg RL, Nelson KG, Peralta-Carcelen M, Johnson SE, DuBard MB et al. Acid-base status at birth and subsequent neurosensory impairment in surviving 500 to 1000 gm infants. *American Journal of Obstetrics & Gynecology* 1994; 170(1 Pt 1):48-53.

(77) van de Riet JE, Vandenbussche FP, Le CS, Keirse MJ. Newborn assessment and long-term adverse outcome: a systematic review. *American Journal of Obstetrics & Gynecology* 1999; 180(4):1024-1029.

(78) Stroup DF, Berlin JA, Morton SC, Olkin I. Meta-analysis of observational studies in epidemiology. *JAMA* 2000; 283(15):2008-2012.

(79) Owen CG, Whincup PH, Odoki K, Gilg JA, Cook DG. Birth weight and blood cholesterol level: a study in adolescents and systematic review. *Pediatrics* 2003; 111(5):1081-1089.

(80) Whincup PH, Kaye SJ, Owen CG, Huxley R, Cook DG, Anazawa S et al Birth weight and risk of type 2 diabetes a systematic review. *JAMA* 2008; 300(24):2886-97.

(81) Bharti B, Bharti S. A review of the Apgar score indicated that contextualization was required within the contemporary perinatal

and neonatal care framework in different settings. *Journal of Clinical Epidemiology* 2005; 58(2):121-129.

(82) Schmidt B, Kirpalani H, Rosenbaum P, Cadman D. Strengths and limitations of the Apgar score: a critical appraisal. *Journal of Clinical Epidemiology* 1988; 41(9):843-850.

(83) ACOG committee opinion. Use and abuse of the Apgar score. Number 174-July 1996 (replaces No. 49, November 1986). Committee on Obstetric Practice and American Academy of Pediatrics: Committee on Fetus and Newborn. American College of Obstetricians and Gynecologists. *International Journal of Gynaecology & Obstetrics* 1996; 54(3):303-305.

(84) Akers J, Aguiar-Ibáñez R, Baba-Akbari Sari A, Beynon S, Booth A. Systematic Reviews: CRD's guidance for undertaking reviews in health care 2009. York UK, NHS Centre for Reviews and Dissemination, University of York.

(85) Khan KS, Dinnes J, Kleijnen J. Systematic reviews to evaluate diagnostic tests. *European Journal of Obstetrics Gynecology and Reproductive Biology* 2001; 95: 6-11.

(86) Deeks J, Khan KS, Song F. Data synthesis. Khan KS, Ter Riet G, Glandville J, editors. Undertaking systematic reviews of research on effectiveness: CRD's guidance for carrying out commisioning reviews. [2nd Ed]. 2001. York (UK), University of York.

(87) Deville W, Buntinx F. Guidelines for conducting systematic reviews of studies evaluating the accuracy of diagnostic studies. Knotterus JA, editor. The evidence base of clinical diagnosis 2002: 145-165. London (UK), BMJ publishing group.

(88) Irwig LM, Tostesen A, Gatsonis C. Guidelines for meta-analysis evaluating diagnostic tests. *Annals of Internal Medicine* 1994; 120: 667-676.

(89) Rutjes AW, Reitsma JB, Di N, Smidt N, van Rijn J, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *Canadian Medical Association Journal* 2006; 174(4):469-476.

(90) Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA,Glasziou PP, Irwig LM. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 2003; 326(7379):41-44.

(91) Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of

studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology* 2003; 3:25.

(92) Whiting P, Harbord R, Kleijnen J. No role for quality scores in systematic reviews of diagnostic accuracy studies. *BMC Medical Research Methodol*ogy 2005; 5:19.

(93) Pepe MS, James H, Longton G, Leisenring W, Newcombe P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *American Journal of Epidemiology* 2004; 159:882-890.

(94) Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. *American Journal of Epidemiology* 1987; 125(5):761-768.

(95) Deeks J. Systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001; 323:157-162.

(96) Higgins JP, Thompson SG, Deeks JJ. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327:557-560.

(97) Higgins JT. Heterogeneity in meta-analysis should be expected and appropriately quantified. *International Journal of Epidemiology* 2008; 37:1158-1160.

(98) Sankey S, Weissfeld L, Fine M, Kapoor W. An assessment of the use of the continuity correction for sparse data in meta-analysis. *Communications in Statistics- Simulation and Computation* 1996; 25:1031-1056.

(99) Freemantle N, Calvert M, Wood J, Eastaugh J, Griffin C. Composite outcomes in randomized trials greater precision but with greater uncertainty? *JAMA* 2003; 289(19):2554-2559.

(100) Deeks J, Higgins JPT, Altman D, (Editors). Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions version 5.0.1. The Cochrane Collaboration; 2008.

(101) Riley RD, Higgins JP, Deeks J.J. The interpretation of random effects meta-analysis. *BMJ* 2011; 342:d549.

(102) Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology* 2005; 58(10):982-990.

(103) Harris RJ, Bradburn M, Deeks J, Harbord RM, Altman D, Steichen T et al. METAN: Stata module for fixed and random effects meta-

analysis. 2006. Statistical Software Components S456798, Boston College Department of Economics, revised 2009.

(104) Harbord RM. METANDI: Stata module to perform meta-analysis of diagnostic accuracy. 2008. Statistical Software Components S456932, Boston College Department of Economics.

(105) Zamora J, Abraira V, Muriel A, Khan KS, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. BMC Medical Research Methodology 2003; 12 (6): 31.

(106) Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *Journal of Clinical Epidemiology* 2005; 58:882-893.

(107) Song F, Khan KS, Dinnes J, Sutton AJ. Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. *International Journal of Epidemiology* 2002; 31:88-95.

(108) Sterne JAC, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011; 342:d4002.

(109) Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias. *JAMA* 2006; 295(6):676-680.

(110) Harbord RM, Harris RJ, Sterne JAC, Steichen T. METABIAS: Stata module to test for small-study effects in meta-analysis. 2009. Boston College Department of Economics, Statistical Software Components S404901.

(111) Heller G, Schnell RR, Misselwitz B, Schmidt S. Umbilical blood pH, Apgar scores, and early neonatal mortality. *Zeitschrift fur Geburtshilfe und Neonatologie* 2003; 207(3):84-89.

(112) Fahey J, King TL. Intrauterine asphyxia: Clinical implications for providers of intrapartum care. *Journal of Midwifery and Women's Health* 2005; 50(6): 498-506.

(113) Milkhu vs NorthWest Hospitals NHS trust 2003 EWHC 94 (QB). 2003. EWHC 94 (QB). http://www.lexisnexis.com/uk/legal/search/runRemoteLink.do?service=citator&csi=279841&remotekey2=%5B2003%5D+All+ER+%28D%29+333+%28Jan%29&remotekey1=REPORTCITATION&risb=21_T8890416457&citatorCC=GB.

(114) Hill AB. The environment and disease. Association or causation? *Proceedings of the Royal Society of Medicine* 1965; 58:295-300.

(115) Weed DL. On the use of causal criteria. *International Journal of Epidemiology* 1997; 26(6):1137-1141.

(116) Weed DL. Interpreting epidemiological evidence: how meta-analysis and causal inference methods are related. *International Journal of Epidemiology* 2000; 29:387-390.

(117) Thorp JA, Rushing RS. Umbilical cord blood gas analysis. *Obstetrics & Gynecology Clinics of North America* 1999; 26(4):695-709.

(118) Ferriero DM. Neonatal brain injury. *New England Journal of Medicine* 2004; 351:1985-1995.

(119) Johnston MV, Trescher WH, Ishida A, Nakajima W. Neurobiology of hypoxic ischaemic injury in the developing brain. *Journal of Child Neurology* 2001; 15:588-591.

(120) Lijmer JG, Bossuyt PM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Statistics in Medicine* 2002; 21(11):1525-1537.

(121) Khan KS, Bachmann LM, ter Riet G. Systematic reviews with individual patient data meta-analysis to evaluate diagnostic tests. *European Journal of Obstetrics Gynecology and Reproductive Biology* 2003; 108(2):121-125.

(122) Blumenthal I. Cerebral palsy- medicolegal aspects. *Journal of the Royal Society of Medicine* 2001; 94:624-626.

(123) Whyte H, Hannah ME, Saigal S, Hannah WJ, Hewson S, Amankwah K et al. Outcomes of children at 2 years after planned cesarean birth versus planned vaginal birth for breech presentation at term: the international randomized term breech trial. *American Journal of Obstetrics & Gynecology* 2004; 191:864-871.

(124) Hemingway H, Riley RD, Altman D. Ten steps towards improving prognosis research. *BMJ* 2009; 339:b4184.

(125) Doctor BA, O'Riordan MA, Kirchner HL, Shah D, Hack M. Perinatal correlates and neonatal outcomes of small for gestational age infants born at term gestation. *American Journal of Obstetrics & Gynecology* 2001; 185(3):652-659.

(126) Ferdynus C, Quantin C, Abrahamowicz M, Platt R, Burguet A, Sagot P et al. Can Birth Weight Standards Based on Healthy Populations Improve the Identification of Small-for-Gestational-Age

Newborns at Risk of Adverse Neonatal Outcomes? *Pediatrics* 2009; 123(2):723-730.

(127) Gardosi J. Intrauterine growth restriction: new standards for assessing adverse outcome. *Best Practice and Research: Clinical Obstetrics and Gynaecology* 2009; 23:741-749.

(128) Walther FJ, Ramaekers LH. The ponderal index as a measure of the nutritional status at birth and its relation to some aspects of neonatal morbidity. *Journal of Perinatal Medicine* 1982; 10(1):42-47.

(129) Botero D, Lifshitz F. Intrauterine growth retardation and long-term effects on growth. *Current Opinion in Pediatrics* 1999; 11(4): 340-347.

(130) Cochrane handbook for systematic reviews of interventions. Higgins JPT, Green S, editors. 2008. The Cochrane Collaboration.

(131) Macaskill P, Gatsonis C, Deeks J, Harbord RM, Takwoingi Y. Analysing and Presenting results. In: Deeks J, Bossuyt P, Gatsonis C, editors. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Version 1.0 ed. The Cochrane Collaboration; 2010.

(132) NCC-WCH. Antenatal care: routine care for the healthy pregnant woman. [Second]. 2009. http://guidance.nice.org.uk/CG62/NICEGuidance/pdf/English

(133) Gardosi J. Ethnic differences in fetal growth. *Ultrasound in Obstetrics and Gynecology* 1995; 6:73-74.

(134) Phillips DIW. Birth weight and adulthood disease and the controversies. *Fetal and Maternal Medicine Review* 2006; 17(3):205-227.

(135) Arcangeli T, Thilaganathan B, Hooper R, Khan KS, Bhide A. Neurodevelopmental delay in small babies at term. A systematic review. *Ultrasound in Obstetrics and Gynecology* 2012; Accepted article(**doi: 10.1002/uog.11112.**).

(136) Knudsen LB, Olsen J. The Danish medical birth registry. *Danish Medical Bulletin* 1998; 45(3):320-323.

(137) Royston P, Sauerbrei W. Multivariable model-building: A pragmatic approach to regression analysis based on fractional polynomials for modelling continuous variables. Chichester: John Wiley; 2008.

(138) Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010; 340:c221.

(139) Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ* 2009; 338:b604.

(140) Apgar V, James LS. Further observations on the newborn scoring system. *American Journal of Disease in Childhood* 1962; 104(4): 419-428.

(141) O'Donnell CPF, Kamlin OF, Davis PG, Carlin JB, Morley CJ. Interobserver variability of the 5-minute Apgar score. *Journal of Pediatrics* 2006; 149:486-489.

(142) Clark DA, Hakanson DO. The inaccuracy of Apgar scoring. *Journal of Perinatol*ogy1988; 8:203-205.

(143) Hegyi T, Carbone T, Anwar M, Ostfeld B, Hiatt M, Koons A et al. The apgar score and its components in the preterm infant. *Pediatrics* 1998; 101(1):77-81.

(144) Casey DM, Tella N, Turesky R, Labrecque M. Therapeutic hypothermia: treatment for hypoxic-ischaemic encephalopathy in the NICU. *Neonatal Network* 2011; 30(6):370-380.

(145) Drury PP, Bennet L, Gunn AJ. Mechanisms of hypothermic neuroprotection. *Seminars in fetal and neonatal medicine* 2010; 15:287-292.

(146) Inder TE, Volpe JJ. Mechanisms of perinatal brain injury. *Seminars in Neonatology* 2000; 5:3-16.

(147) Westin B. Hypothermia in the resuscitation of the neonate: a glance in my rear-view mirror. *Acta Paediatrica* 2006; 95:1172-1174.

(148) Mann TP, Elliot RIK. Neonatal cold injury due to accidental exposure to cold. *Lancet* 1957; 272:229-234.

(149) Silverman WS, Fertig JW, Berger AP. The influence of the thermal enviroment on the survival of newly born premature infants. *Pediatrics* 1958; 22:876-885.

(150) Jacobs S, Hunt R, Tarnow MW, Inder T, Davis P. Cooling for newborns with hypoxic ischaemic encephalopathy. *Cochrane Database of Systematic Review* 2008;(4).

(151) Gunn AJ, Gluckman PD, Gunn TR. Selective head cooling in newborn infants after perinatal asphyxia: a safety study.[see comment]. *Pediatrics* 1998; 102(4):885-892.

(152) Akisu M, Huseyinov A, Yalaz M, Cetin H, Kultursay N. Selective head cooling with hypothermia suppresses the generation of platelet activating factor in cerebrospinal fluid of newborn infants with perinatal asphyxia. *Prostaglandins, Leukotrienes and Essential Fatty Acids* 2003; 69:45-50.

(153) Gluckman PD, Wyatt J, Azzopardi D, Ballard R, Edwards D, Ferriero D. Selective head cooling with mild systemic hypothermia after neonatal encephalopathy: multicentre randomised controlled trial. *Lancet* 2005; 365:663-670.

(154) Lin Z-L, Yu HM, Lin J, Chen SQ, Liang ZQ, Zhang ZY. Mild hypothermia via selective head cooling as neuroprotective therapy in term neonates with perinatal asphyxia: An experience from a single neonatal intensive care unit. *Journal of Perinatology* 2006; 26(3): 180-4.

(155) Inder T, Hunt R, Morley C, Gunn TR, Gluckman PD, Gunn AJ. Randomised trial of systemic hypothermia selectively protects the cortex on MRI in term hypoxic-ischaemic encephalopathy. *Journal of Pediatrics* 2004; 145:835-837.

(156) Shankaran S, Laptook AR, Ehrenkranz R, Tyson J, McDonald S, Donovan E. Whole-body hypothermia for neonates with hypoxic, ischaemic encephalopathy. *New England Journal of Medicine* 2005; 353:1574-1584.

(157) Eicher D, Wagner C, Katikaneni L, Hulsey T, Bass T, Kaufman D et al. Moderate hypothermia in neonatal encephalopathy: Efficacy outcomes. *Pediatric Neurology* 2005; 32:11-17.

(158) Shankaran S, Laptook A, Ehrenkranz R, Donovan E, Fanaroff AA, et al. Whole body hypothermia for neonates with hypoxic ischaemic encephalopathy. *Pediatrics* 2002; 110:377-385.

(159) Azzopardi D, Strohm B, Edwards DA, Dyet L, Halliday HL, Juzczak E. Moderate hypothermia to treat perinatal asphyxial encephalopathy. *New England Journal of Medicine* 2009; 361(14):1349-1358.

(160) Edwards AD, Brocklehurst P, Gunn AJ, Halliday HL, Juzczak E, Levene M et al. Neurological outcomes at 18 months of age after moderate hypothermia for perinatal hypoxic ischaemic encephalopathy: synthesis and meta-analysis of trial data. *BMJ* 2010; 340:c363.

(161) Simbruner G, Mittal RA, Rohlmann F, Muche R, et al. Systemic hypothermia after neonatal encephalopathy: outcomes of neo.neuro.network RCT. *Pediatrics* 2010; 126:e771-e778.

(162) Jacobs S, Morley CJ, Inder TE, Stewart MJ, Smith KR, McNemara PJ et al. Whole body hypothermia for term and near-term newborns with hypoxic ischaemic encephalopathy. *Archives of Pediatrics and Adolescent Medicine* 2011; 165(8):692-700.

(163) Regier DA, Petrou S, Henderson J, Eddama O, Patel N, Strohm B et al. Cost-effectiveness of therapeutic hypothermia to treat neonatal encephalopathy. *Value in Health* 2010; 13(6):695-702.

(164) Kirpalani H, Barks J, Thorlund K, Guyatt G. Cooling for neonatal hypoxic ischemic encephalopathy: do we have the answer? *Pediatrics* 2007; 120(5):1126-1130.

(165) UK Toby Cooling Register Clinicians Handbook. 2010. www.npeu.ox.ac.uk/tobyregister

(166) Malin GL, Morris RK, Khan KS. Strength of association between umbilical cord pH and perinatal and long term outcomes: systematic review and meta-analysis. *BMJ* 2010; c1471.

(167) Barton P, Bryan S, Robinson S. Modelling in the economic evaluarion of health care: selecting the appropriate approach. *Journal Health Services and Research Policy* 2004; 9:110-118.

(168) Data Treeage [Williamstown, MA: Treeage Software Inc; 2005.

(169) Surveillance of cerebral palsy in Europe. Surveillance of cerebral palsy in Europe: a collaboration of cerebral palsy surveys and registers. *Developmental medicine and Child Neurology* 2000; 42(12):816-824.

(170) DoH payment by results team. 2010-2011 reference costs publication. 2011. www.dh.gov.uk/nhscosting

(171) Laudicella M, Rosen KO, Street A. What explains the variation in the cost of treating patients in English obstetric specialties. 2009. University of York, Centre for Health Economics. www.york.ac.uk/inst/che/pubs

(172) ISD Scotland. Childhood hospital admissions and mortality. 2008. http://data.gov.uk/dataset/childhoodhospitaladmissionsandmortality

(173) Hippisley-Cox J, Vinogradova Y. Trends in consultation rates in general practice 1995/1996 to 2008/2009: Analysis of the QResearch database. 2009. www.ic.nhs.uk

(174) Unit costs of health and social care. Curtis L, editor. 2011. The University of Kent, Personal social Services Research Unit. www.pssru.ac.uk

(175) Sanderson D, Wright D, Acton C, Duree D. Cost analysis of child health surveillance. *Health Technology Assessment 2001* 2001; 5(36).

(176) Briggs A, Claxton K, Sculpher M. Decision modelling for health economic evaluation. 1 ed. Oxford: Oxford University press; 2006.

(177) National Institute for Clinical Excellence. Guide to the methods of health technology appraisal. 2004. http://www.nice.org.uk/niceMedia/pdf/TAP_Methods.pdf

(178) Centre for Evidence Based Medicine. Likelihood ratios. 2012. http://www.cebm.net/index.aspx?o=1043